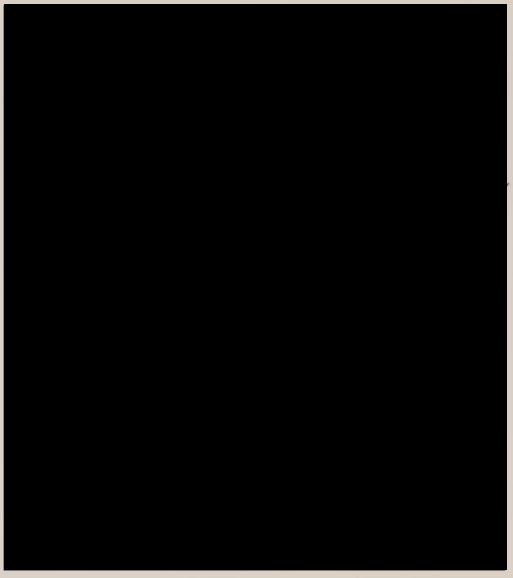
STUDIES IN SEISMIC SCATTERING

APPROVED BY
DISSERTATION COMMITTEE:



I can't see myself as a professional student--in school till I'm 26 . . .
--Steve Cardimona (Junior year of high school, spring, 1980)

I dedicate this dissertation to my parents,
since this may be the only book I ever write!

STUDIES IN SEISMIC SCATTERING

by

STEVEN JAMES CARDIMONA, B.S.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 1992

PREFACE

I thank Jan Garmany and all the members of my supervisory committee for their suggestions and support during this research, especially when I tended to be a bit too independent (i.e., I didn't talk with them enough!). With six years of one's life lost at a single institution for higher education, the question must arise as to what has been gained. If it was up to me to bring fame to my supervisor or glory to my supervising committee members, then I have failed miserably. Nor have I solved any burning scientific question of my own, and I am still pondering the pressing problem I had six years ago: "What is the connection between squinting and sneezing? That is, why does sunlight, or a strong incandescent light (not fluorescent, mind you!) cause many people to sneeze?" Of course, the fact that my graduate research in geophysics has not answered this question is not a failing.

It is true that I have learned a lot. In fact, one of my first major lessons came in applying for a National Science Foundation graduate fellowship. In the summary of the distribution of fellowship offers for 1987, it was very enlightening to see that students of the Earth Sciences received the lowest percentage of fellowships; even below the number offered to students of psychology! Perhaps in that field I would have been able to study the squinting/sneezing problem.

A glance at some common literature will show that Ph.D.'s in the social sciences such as archeology or anthropology (or psychology?!) are most often referred to as 'scholars'. Persons with Ph.D.'s in the physical sciences are referred to as 'researchers' or 'scientists'. This is fine; but, I feel sure that, while the use of 'scholar' is fairly exclusive, the same common literature will use 'researchers' or 'scientists' to describe anyone with a B.S., an M.S. or a Ph.D. working in the laboratory or out in the field studying physical processes.

If not a scholar, what am I? My wife is quick to point out that she has completed far more studies in philosophy than I, yet I am a candidate for the degree of Doctor of Philosophy! Nevertheless, it has been rewarding working on problems in geophysics, and I am personally much better for the time I spent working toward my Ph.D. degree. Alas, as simple as that may seem, that must be enough.

In my time at the University of Texas and the Institute for Geophysics, I have acquired quite a list of people and organizations for which I would like to give thanks. First, to the UTIG Fellowship Committee of 1986-87, for seeing in me a person worthy of work at Austin. I want to thank the University of Texas Department of Geological Sciences Admissions and Support Committee as well, for awarding me fellowship money from industry grants, in particular from Arco and Amoco. Indeed, thanks go to the Geology Foundation of UT as well, for administering the fellowship awards and for offering Professional Development scholarships, without which I would not have been as free to present my work at international meetings. I want to thank Art Maxwell of the UTIG, who helped me through a couple low income times with Institute money from some account someplace (I don't care where!). Special thanks go to Paul Stoffa who, despite having a gaggle of students of his own, watched out for my interests when he had the chance.

I do not want to limit my thanks to only those tied in with giving me money. I want to thank my brother Joe for keeping the faith, my brother Dave for blazing the trail, my brother Tom for making the music, and Dave Gableman for all those games, games, games!. I thank Scott Davis for helping me go for the guru and Barbara Wahl for enlightening me on the subject of omega to the omega (it's clear as mud to me now!) My thanks to Cathy Mayes for introducing me to the Court Jester; Seiichi Nagihara for bowling in the morning at half price; Jinyong Oh for sharing all our problems; Mark Shorey for my first international correspondence; and Denise Kakas for shooting all those slides. I want to thank Cliff Frohlich for always offering advice from his man-of-the-world (i.e., twisted) point-of-view. A special thanks goes to Hugh Winkler for not pounding me senseless when I stole his desk. It's been fun, and I would do it all again if I had the chance.

This research was partially supported by grants from the National Science Foundation and the Advanced Research Program sponsored by the Texas Higher Education Coordinating Board.

STUDIES IN SEISMIC SCATTERING

Publication No._____

Steven James Cardimona, Ph.D.
The University of Texas at Austin, 1992

Supervisor: Clark R. Wilson

Modern methods of seismic data analysis are tending to inversion through model fitting, i.e., actually finding the best model of the Earth's subsurface which would produce the amplitude and phase variation in the observed data. An understanding of seismic scattering is fundamental to this form of data analysis. This dissertation involves the study of seismic scattering and its use in the inverse problem, applying full waveform inversion ideas in novel situations. The terminology and methodology of inverse theory may sometimes hide what is going on, and may make it difficult to connect the results with those from more familiar techniques. In Chapter 1 I show that, with the appropriate choice for the model parameters, the first iteration of the nonlinear least-squares seismic waveform inversion algorithm reduces to classical results from linear filter theory.

I use the idea of the adjoint of the Frechet derivative linear operator in Chapter 2 to understand smoothing in the waveform inversion, which manifests itself as a new sensitivity function incorporating the smoothing information. This gives us physical intuition into the wave equation based inverse problem. My mathematical analysis is general; however, using sensitivity functions for the paraxial equation in ray centered coordinates, I show a specific application to full

waveform imaging in a tomographic experiment where only phase information (travel-time data) is normally used.

I consider an inversion of teleseismic data from some deep earthquakes in Chapter 3. I use the phase and amplitude variation in the seismic signals in an imaging technique derived from inverse theory and digital signal analysis, interpreting the coherent energy in the coda of the first arrival as due to scattering from upper mantle discontinuities. Applying an inversion through iterative forward modeling, I measure the depth variation of the spinel-perovskite upper mantle phase transition within the subduction zone region. This measurement allows me to characterize the variation of the transition with respect to pressure and temperature. My results are consistent with convection in a model of a chemically homogeneous mantle, where the presence of the phase transition at around 670 km depth disrupts the full mantle convection patterns.

TABLE OF CONTENTS

Getting back on track is hard to do when you don't know where you are going.
--Beverly Urton (1987)

INTRODUCTION TO THIS DISSERTATION

Scattering

This dissertation concerns the scattering of seismic waves. Seismic scattering implies the propagation of wavefields in an elastic Earth, with conversions between the different types of waves. These wavefields consist of body waves which travel throughout the Earth, and surface waves which travel along the surface of the Earth. Of the body waves, there are two types, compressional waves and shear waves. Elastic waves are in general vector waves, needing an amplitude with a direction (a vector) of particle motion associated with it to describe the wave.

The mathematics contained within these pages relates to acoustic scattering and scalar wave phenomenon. The use of the term 'acoustic' waves will often imply 'sound' waves which are compressional waves; however, I consider the conversion of shear waves to compressional waves as well as simply compressional wave scattering, so I use the term seismic scattering to emphasize my use of the shear wave information in an elastic Earth. My analysis treats the shear waves as scalar waves, however, and I do not deal with elastic vector wave scattering.

I define 'scattering' to mean anything but the incident (the source) wave. This definition comes from ideas in diffraction theory, which seeks to describe wave propagation in heterogeneous or obstructed media. In the first two appendices at the end of this dissertation, I review the basic concepts from scattering theory and diffraction theory. The term diffraction is used often to describe wave scattering from a single point in the material, and this is consistent with diffraction theory. Wave reflection and transmission result from the constructive interference of diffractions from a number of point scatterers

which define the material discontinuity. With my definition of scattering, the total wavefield u is a sum of the incident wave and the scattered wavefield

$$u = u_i + u_s \ . \qquad (1)$$

The concept of the incident wave can be generalized to that of an incident wavefield. The incident wavefield would contain all the wave interactions with some current Earth model, leaving the scattered wavefield to be that part of the data which is unaccounted for in the forward theory. For example given a surface seismic reflection experiment with a layer-cake model for the subsurface geology, the incident wavefield could be taken as the direct wave along with reflections and multiples from known discontinuities in the model. The scattered wavefield is that which is unaccounted for in the model, i.e. reflections from interfaces not in the current model, diffractions from lateral heterogeneity, shear-wave splitting due to anisotropy, surface waves, etc.

If the incident wave propagates through a homogeneous isotropic medium, it will not be disrupted in any way. Scattering is a result of interaction of the wave with heterogeneities or obstructions within the medium through which the wave is propagating. These heterogeneities are described by a variation in velocity of the wave or density of the medium, most often due to chemical differences, i.e., physically different types of materials. Earth mineral assemblages go through phase transformations related to the existing temperature and pressure. These phase transformations can introduce seismic discontinuities within a chemically homogeneous material. It is exactly the fact that waves scatter due to interactions with the heterogeneity that allows us to use the scattering to understand the variation in properties of the medium.

When the scattered energy is reflected by the material heterogeneity so that it travels in generally the opposite direction as the incident wave, we get what is called back-scattering. In Chapter 1 of this dissertation I study reflections and the reflectivity function for a one dimensional Earth model consisting of plane layers of different materials in a 'layer-cake' structure. The example application in Chapter 2 deals with solutions to the paraxial, or one-

way wave equation when the homogeneous background model is perturbed with some material heterogeneity. In that chapter I focus on only forward-scattered energy, where the scattered waves travel in basically the same direction as the incident wave. The analysis of the earthquake seismic records in Chapter 3 considers the possibility of either forward or back-scattered waves.

Seismic data

In defining scattering as described above, it must include the parts of the recorded seismic data that we are interested in, as well as much of the data that we are not interested in. The former we want to interpret as 'signal', and the latter we call 'noise'. The physical problem that we are studying will define what the signal is that we will look for in the data. For example, in a land seismic exploration reflection experiment with a buried explosive source, the reflections (coherent back-scattering) of waves from deeper Earth structure is the signal we look for in the recorded data. With this kind of experimental geometry, a wave propagating up from the buried source, reflecting off of the surface of the Earth and propagating back into the medium becomes the surface 'ghost', and is noise in the data unless addressed directly. In a global earthquake seismic study using deep earthquakes, such as in Chapter 3 of this dissertation, it is this same free-surface reflection, e.g. the pP phase, which we look for as signal to determine the depth of the earthquake.

The theoretical motivation in Chapter 2 addresses this understanding of signal versus noise, generalizing the concept in terms of smoothing functions in the solution to the forward problem. This is important since it is these smoothing functions which enter into the inversion of the data as damping operators. The 'forward problem' is not to be confused with 'forward-scattering' described above. The forward problem is the analytical or numerical solution to the physical system of interest, giving the scattering result (forward and/or back-scattering) due to some Earth model. In Chapter 3, I consider three separate analyses of the data in which I depend upon my different definitions of signal versus noise. For example, when looking for back-scattered

compressional energy, any forward-scattered shear to compressional wave conversions are noise in the data records.

One has to be careful when speaking of the seismic frequencies of interest in any particular study, because what is high frequency to global earthquake seismologists is low frequency to seismic explorationists. The theoretical analyses in Chapters 1-3 are general and can be applied to seismic data of any frequency band. Chapter 3 deals with use of high frequency global earthquake data around .5 Hz (2 second period), which is one of the important aspects of the analysis. It is the shorter period global earthquake data which can give us information on the lateral variation in the long wavelength Earth structure. The frequencies of the earthquake data in Chapter 3 are very low, however, in comparison to studies of shallow Earth structure in an exploration seismic geometry such as the example application in Chapter 2, where frequencies between 10 and 60 Hz are of interest.

Typically, we make a 'high frequency approximation' when we can, in order for the mathematical analysis to be simplified. When we do this, we can describe wave solutions in terms of base characteristics instead of full characteristics. In geometrical optics, the base characteristics are the rays of propagation. This is ray theory, and we can apply it to seismic wave propagation. In general, the physical system of interest defines a 'typical' dimension L, and the frequency content of the signal defines a 'typical' wavelength Ω in the material. Then

$$\frac{L}{\Omega} >> 1 \qquad (2)$$

gives a general condition for being able to apply the high frequency approximation.

Earth models

Luckily, many of the effects in seismic data can be described with a linear convolution model, and can be studied with linear filter theory. Given a

source-time function s(t) and a receiver instrument response I(t), the data d(t) can be described as a convolution of the source and instrument with the Earth response G(t)

$$d(t) = s(t) * G(t) * I(t) , \tag{3}$$

where the convolution operation '*' is defined by, for time functions a(t) and b(t),

$$c(t) = a(t) * b(t) = \int_{-\infty}^{\infty} a(\tau)\, b(t - \tau)\, d\tau . \tag{4}$$

Expanding the Earth response to show near source response S(t), path effects P(t) and near receiver response R(t) gives

$$G(t) = S(t) * P(t) * R(t) . \tag{5}$$

From equation 4, it is clear that the convolution operation is commutative and associative. Thus if I am interested in the near source term, I can rewrite equation 3 using equation 5 to get

$$d(t) = S(t) * W(t) , \tag{6}$$

where

$$W(t) = s(t) * P(t) * R(t) * I(t) . \tag{7}$$

In equation 6 I have associated everything I am not interested in with the function W(t). I address seismic waveform inversion with the convolutional model in Chapter 1, and I use the theoretical ideas and filtering operations in

Chapter 3 to estimate a W(t) function so that I can concentrate my analysis on near source structure.

Inverse theory

Once we have a model of the Earth which describes that part of the data we call signal, the problem of inversion is to take that signal and invert it for the specific parameters of the model. The 'data space' is made up of the signal we are interested in plus any extra noise. The 'model space' is the space of all possible vectors of model parameters. The inverse problem is solved in general through iterative forward modeling, where synthetics created for a given model are compared directly with the data in some way. The 'image space' is a resolving space for each model parameter during each iteration. When the iterative process ends, we have some model parameters which create synthetics that best compare with the data, hence best represent the signal in the data.

We can make the comparison between the data d(t) and the synthetics a(t) derived for a given model in a variety of ways, in each case evaluating an objective function which we wish to minimize or maximize. A normalized error energy function which we would try to minimize might look like

$$E = \frac{\sum_i (d_i - a_i)^2}{\sum_i d_i d_i} \quad . \tag{8}$$

In equation 8, the total error energy (sum of the squared error) is normalized by the total energy in the data. We might try to maximize a normalized crosscorrelation function given by

$$C = \frac{\sum_i d_i a_i}{\sum_i d_i d_i} \quad . \tag{9}$$

Equation 9 defines a function giving the energy that is in phase between the data and the synthetic, again normalized here by the total energy in the data itself. Maximizing the coherence between the data and the synthetic would involve the use of some measure related to the semblance as the objective function. The semblance gives the ratio of the total energy in the stack of the data with the synthetic to the sum of the energy of the data and the synthetic separately. This function might look like

$$S = \frac{\sum_i (d_i + a_i)^2}{2 \sum_i d_i^2 + a_i^2} \quad . \tag{10}$$

Often the model space is very large, so testing every model parameter vector may be impossible. Searching for the best vector of model parameters can be done in a variety of ways. Monte Carlo methods consist of a random search throughout model space, in which it is hoped the best model can be found without testing all of the models. Simulated annealing and genetic algorithms attempt to search through model space rapidly to obtain the best global model vector, borrowing ideas from natural systems to direct the search. Calculus-based methods offer a locally guided search throughout model space.

In calculus-based techniques, the search guidance comes from the local linearization of the forward problem for scattering. I review the basic analysis technique for this method in Appendix III, and I use these ideas in applications to full waveform data inversion in Chapters 1-3. In Chapter 1 I take a linear model and use the local linearization to get my result in one iteration. This one-step calculation is an 'imaging', or estimation of the model parameters. In the example application of Chapter 2 I image perturbations to a background slowness model in a cross bore-hole transmission problem. In Chapter 3, I use the same ideas to image the variation in shear wave to compressional wave transmission coefficients in the upper mantle. From the imaging results, I chose a model of a single perturbed discontinuity to describe the 670 km

seismic discontinuity. The perturbation in the interface is a downwarping within the descending lithospheric slab, described by the model parameters of slab thickness and depth change. With only two model parameters, and physical limitations on how much each one can vary, I search through the model space directly to find the best model parameters which explain the data imaging results, relative to objective functions such as in equations 8-10.

I have focused the mathematics in this dissertation on solutions to the forward problem of seismic scattering in order to help interpret the results from application of a calculus-based inversion. For the inverse theory, I only reproduce the important aspects needed for the subjects of my chapters. There is a bit of overlap of material in the theory sections of Chapters 1-3 and the appendices. This is in order to introduce that part of the theoretical work most pertinent to the topic of each chapter. I wait until Chapter 2 to develop most of the background for general scattering problems, so that in Chapter 1 I may concentrate on a simple model and the intuitive results of a simple inversion calculation.

Damn the torpedoes, full speed ahead!

--Admiral Farragut (1864)

CHAPTER 1: WAVEFORM INVERSION AND DIGITAL FILTER THEORY

INTRODUCTION

For many years, researchers have used seismic data processing techniques such as digital filter theory to enhance the signal-to-noise ratio of the observed data in order to obtain the best possible image of the subsurface. With our increased knowledge and the increasingly sophisticated computers at our disposal, we can try to find the best model of the subsurface which would produce the observed data by applying an inversion through model fitting.

One method which many researchers are studying is a calculus-based technique for nonlinear waveform inversion, subject to the least-squares criterion (Tarantola and Valette, 1982; Tarantola, 1984, 1986). In this method, the nonlinear problem is attacked through an iterative solution of locally linear problems. Tarantola (1984) has shown how a waveform inversion gradient method is equivalent to a sophisticated migration of the data when the model perturbations are chosen to be changes in velocity values of the medium.

In this chapter I show that, with the appropriate choice for the model parameters, the first iteration of the nonlinear least-squares waveform inversion algorithm reduces to linear filter theory. Before the advent of waveform inversion techniques, Turin (1957) achieved similar results using an information theory analysis. In particular, when the source wavelet is known and the model consists of the amplitudes in a reflectivity time series, the full waveform gradient solution, or steepest descent method, yields a scaled version of the matched filter result (Treitel and Robinson, 1969), while the Newton method solution is proportional to the zero-delay spike deconvolution result (Peacock and Treitel, 1969; Lines and Treitel, 1984).

INVERSE THEORY

Given the nonlinear relationship $\mathbf{u}=g(\mathbf{m})$ between the observed seismic data $\mathbf{u}$ and the model parameters $\mathbf{m}$, many authors (Menke, 1984; Tarantola, 1987) have outlined methods for obtaining solutions to the seismic inverse problem. In the spirit of Backus and Gilbert (1968,1970), and Tarantola and Valette (1982), an iterative least-squares solution to the inverse problem, for iteration (n+1), is given by

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \mu_n \mathbf{T}_n \gamma_n , \tag{1}$$

where $\mathbf{m}$ is the vector of model parameters, and γ is the direction of steepest ascent at $\mathbf{m}_n$. If we let $\mu_n \mathbf{T}_n = \mathbf{H}_n^{-1}$, where $\mathbf{H}$ is the model space curvature operator, then we have the Newton algorithm. Letting $\mathbf{T}_n = \mathbf{I}$ gives the steepest descent method with $\mu_n > 0$ as the step-size parameter. Considering the Newton method (Tarantola, 1987), we have that

$$\delta\mathbf{m}_{n+1} = \mathbf{m}_{n+1} - \mathbf{m}_n = \left(\mathbf{C}_m \mathbf{G}_n^t \mathbf{C}_u^{-1} \mathbf{G}_n + \mathbf{I} \right)^{-1} \left[\mathbf{C}_m \mathbf{G}_n^t \mathbf{C}_u^{-1} \delta\mathbf{u}_n - \left(\mathbf{m}_n - \mathbf{m}_{n-1} \right) \right] . \tag{2}$$

Here, $\delta\mathbf{u}_n = \mathbf{u} - g(\mathbf{m}_n)$ is the current data residual; $g(\mathbf{m}_n)$ is the response to the model for the n^{th} model iteration; $\mathbf{m}_o = \mathbf{m}_{-1}$ is the initial model (so that for the first iteration the gradient part in brackets is just $\mathbf{C}_m \mathbf{G}_o^t \mathbf{C}_u^{-1} \delta\mathbf{u}_o$); $\mathbf{G}$ is the linear operator (Frechet derivative) which maps perturbations in the model space to perturbations in the data space; $\mathbf{G}^t$ is the transpose of $\mathbf{G}$; and $\mathbf{C}_u$ and $\mathbf{C}_m$ are the least-squares functional weighting operators, the data and model covariance matrices, respectively (Tarantola, 1986). The Frechet derivative kernel maps a function of a continuous spatial coordinate to a function of discrete source and receiver locations and time. The kernel for the transpose operator is the same as that of the forward operator except that the variables are said to be transposed, and the mapping is from functions of source and receiver locations and time to functions of space. Depending upon whether we include $\mathbf{C}_m$ or $\mathbf{C}_u$, and, if so, whether they are proportional to the identity operator, we can get various special cases of the above

solution (Wiggins, 1972; Menke, 1984). In particular, if we want to minimize a weighted combination of prediction error and model covariance size, and we let $\mathbf{C}_u=\sigma_u^2\mathbf{I}$ and $\mathbf{C}_m=\sigma_m^2\mathbf{I}$, the first iteration (n=0) of equation 2 for the Newton algorithm gives the damped least-squares solution in the form

$$\mathbf{m}_1 = \mathbf{m}_o + \left(\mathbf{G}_o^t\mathbf{G}_o + \frac{\sigma_u^2}{\sigma_m^2}\mathbf{I} \right)^{-1} \mathbf{G}_o^t \left[\mathbf{u} - g(\mathbf{m}_o) \right] . \tag{3}$$

Equation 3 is the usual starting place for a least-squares data inversion (Lines and Treitel, 1984).

LINEAR CONVOLUTION MODEL

Let us consider a single data trace $u_o(t)$ as the result of a convolution in time of the known source wavelet s(t) with a reflectivity series r(t) given by

$$r(t) = \sum_i r_i \, \delta(t-\tau_i) , \tag{4}$$

so that

$$u_o(t) = \sum_i r_i \, s(t-\tau_i) . \tag{5}$$

We could take both the reflectivity amplitudes and their respective delay times as model parameters. If we choose the delay times to be part of our model space, this would be a truly nonlinear problem requiring a linearized iterative solution. We instead choose to take only the amplitudes r_i as our model parameters. Ignoring diffractions and amplitude corrections for such effects as attenuation, spherical divergence and transmission loss, the model of a reflectivity containing only primaries and no significant multiples allows the reflectivity series to be represented by a linear model (Oldenburg, 1981). Then the first iteration in the nonlinear inversion algorithm should reduce to linear theory. After Goupillaud

(1961), I choose the subsurface layer thicknesses Δz_i such that, for velocity c_i in the i^{th} layer,

$$\frac{\Delta z_i}{c_i} = \Delta t_i = \Delta t \ , \tag{6}$$

where Δt is a constant. We can then represent the reflectivity values as $r_i=r(\tau_i)$, where $\tau_i=i\Delta t$ (i=0,n). For clarity, let us keep time 't' as a continuous variable, even though we have introduced discrete time for the delay times ' τ_i'. Considering perturbations to the reflectivities, it is easy to see from equation 5 that we get

$$\delta u(t) = u(t) - u_o(t) = \sum_i \delta r_i \, s(t-\tau_i) \ . \tag{7}$$

Thus, the Frechet derivative operator acting on perturbations to the model parameter r_i is just the source wavelet with the appropriate delay τ_i. So, for the linear forward problem $\delta\mathbf{u}=\mathbf{G}\delta\mathbf{m}$, we get

$$\delta u(t) = \left[s(t-\tau_o) \ . \ . \ . \ s(t-\tau_n) \right] \begin{bmatrix} \delta r_o \\ . \\ . \\ . \\ \delta r_n \end{bmatrix} \ . \tag{8}$$

The gradient solution

Refering to equations 1 and 2, the first iteration of a full waveform gradient solution for the model perturbations (with $\mu_o=1$ and $\mathbf{G}_o=\mathbf{G}$) is given by

$$\delta\mathbf{m}_1 = \mathbf{C}_m \mathbf{G}^t \mathbf{C}_u^{-1} \delta u(t) \ . \tag{9}$$

If we take $\mathbf{C_m}= \mathbf{C_u}=\mathbf{I}$, let each model parameter update be given by $\delta m_1=\delta r_i=\delta r(\tau_i)=r(\tau_i)$, and the data residual $\delta u(t)=u(t)$ for a first iteration result, we then have

$$r(\tau_i) = \frac{1}{T}\int_0^T dt\, s(t-\tau_i)\, u(t)\ . \qquad (10)$$

Now, τ_i are vertical delay times related to each model parameter r_i. We see from this simple case that the action of the transpose operator $\mathbf{G^t}$ is to 'forward-propagate' the source wavelet and 'back-propagate' the data to the depth related to the delay time τ_i. Here, the two fields are cross-correlated to yield an image density which is proportional to the required change in the model parameter r_i. We also see that the operation on u(t) in equation 10 is that of a cross-correlation, or matched filter of the data u(t) for the known signal s(t) (Treitel and Robinson, 1969). I have scaled the transpose operator by the cross-correlation time window length T in equation 10. Any covariance matrices $\mathbf{C_m}$ and $\mathbf{C_u}$ that are proportional to the identity matrix would simply introduce another scaling into the filter operation. Thus, the first iteration of the gradient solution to the waveform inversion problem yields a matched filter.

The Newton method solution

Using equation 2, the first iteration in a Newton method solution for the model perturbations (with $\mathbf{G_o}=\mathbf{G}$) is given by

$$\delta\mathbf{m}_1 = \left(\mathbf{C_m G^t C_u^{-1} G} + \mathbf{I}\right)^{-1}\left[\mathbf{C_m G^t C_u^{-1}}\ \delta u(t)\right]\ . \qquad (11)$$

If we let $\mathbf{C_m}=\sigma_r^2\mathbf{I}$, and $\mathbf{C_u}= \sigma_u^2\mathbf{I}$, then we have the damped least-squares solution of equation 3, or Marquardt-Levenberg method (Lines and Treitel, 1984). As in equation 10 for the gradient solution, let $\delta m_1=\delta r(\tau_i)=r(\tau_i)$ and $\delta u(t)=u(t)$ for a first iteration result, so that now our solution is given by

$$r(\tau_i) = \left(\mathbf{G}^t\mathbf{G} + \frac{\sigma_u^2}{\sigma_r^2}\mathbf{I} \right)^{-1} \mathbf{G}^t\, u(t)\,, \tag{12}$$

where, for k and l time indices between 0 and n,

$$\left[\mathbf{G}^t\mathbf{G}\right]_{k,l} = \frac{1}{T}\int_0^T dt\; s(t)\, s(t - |k-l|\Delta t) \;=\; a_{|k-l|}\,. \tag{13}$$

Hence, $\mathbf{G}^t\mathbf{G}$ is the Toeplitz auto-correlation matrix for the known wavelet s(t) (Peacock and Treitel, 1969). We have already seen in equation 10 that the operation of $\mathbf{G}^t$ on the data resulted in cross correlating the known source wavelet with the data. Now we have premultiplied that result with the inverse of the auto-correlation matrix for the source wavelet modified by the addition of $\beta\mathbf{I}$, where β (= σ_u^2/σ_r^2) is the Marquardt damping factor (Lines and Treitel, 1984), known in engineering as the prewhitening parameter. With equations 10, 12 and 13, we get the system of normal equations to be solved for the least-squares, zero-delay inverse (Wiener) filtering of the data with respect to the known source wavelet s(t),

$$r(\tau_i) = \left\{ \begin{bmatrix} a_o & a_1 & \cdots \\ a_1 & a_o & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} + \beta\,\mathbf{I} \right\}^{-1} \frac{1}{T}\int_0^T dt\; s(t-\tau_i)\; u(t)\,. \tag{14}$$

Just as seismic source wavelet deconvolution can be studied in terms of the linear inverse theory of Backus and Gilbert (Backus and Gilbert, 1968, 1970; Oldenburg, 1981; Treitel and Lines, 1982), here we see that the first iteration in the Newton method solution to the full waveform inverse problem results in a spike deconvolution of the data. As we increase the value of the Marquardt damping

factor β, which means adding more white noise to the auto-correlation matrix, the method becomes more like a scaled version of the gradient solution in equation 10 (Lines and Treitel, 1984). So we see that the spike deconvolution algorithm approaches matched filtering as we increase the white noise (increase the ratio $\sigma_u^2/\sigma_r^2=\beta$). Increasing the prewhitening parameter reduces the resolution of the result, but helps to stabilize the inversion, and the matched filter result is the best filter in the presence of abundant random noise (Treitel and Robinson, 1969).

CONCLUSIONS

The preceding analysis shows that the first iteration in an iterative least-squares solution to the nonlinear waveform seismic inverse problem reduces to linear digital filter theory when the seismic source wavelet is known and the model is taken as a reflectivity amplitude time series. The gradient solution to the waveform inversion problem yields a matched filter result. The Newton method solution to the waveform inversion problem, where we incorporate the Hessian (model space curvature information) as well as the gradient (steepest descent direction information), results in a spike deconvolution of the data. These results may not be surprising to inversion specialists, but it is gratifying to see that the relationship between the two theories is exactly what we expect. This relationship is not readily seen, since the full waveform inverse theory has used its own notation and approach which in many ways are different from that of digital filter theory. Recognizing the relationship between waveform inversion and filter theory for a simple model may help in the intuitive understanding of inversion for more elaborate models of the Earth. The gradient solution handles the phase information in the data through forward-propagation of the source and back-propagation of the data residual. It is the model space curvature operator which works more on the amplitude and increases the resolution of the inversion output. In filter theory, if our data is band-limited or has holes in the spectrum, we know that we cannot recover absent frequencies through deconvolution. So too the resolution of any waveform inversion result will still be limited by the spectral content of the data. At each step of the inversion, the resolution of the model parameters takes place in an image space, and is achieved through a cross-correlation operation on the data.

We were made fit for what we must do by what we had done.
--Arthur Pendragon (from *Firelord*, by Parke Godwin)

CHAPTER 2: A SENSITIVE EXTENSION OF RAY THEORY

INTRODUCTION

Ray trace tomography has seen enormous popularity in seismology in recent years, offering scholars a way to invert seismic travel times for an estimate of the background velocities within the propagating medium. Stork and Clayton (1991), for example, give a recent analysis and review of the methodology of ray trace tomography. Lines and LaFehr (1989) apply a straight-ray tomographic analysis, then check their results using more elaborate ray-tracing and finite difference modeling. Bregman, et al. (1989), depart from using just straight rays in their tomography problem. They still use an SVD analysis to get the generalized inverse, and they apply a damped least squares approach with constant covariance matrices that are proportional to the identity. Much current work has gone into studying ray theory and producing efficient calculations of travel times for use in a tomography problem (Vidale, 1988, 1990; Bortfeld, 1989; Moser, 1991; van Trier and Symes, 1991). Travel time information, however, can only give smooth variations in velocities. To image and invert for discontinuities in the velocity function, researchers need to use amplitude information in the seismic signal, and more than just forward-scattered, first arrival traveltime information. Tomographers are working toward bridging the gap between traveltime tomography and full waveform inversion. Diffraction tomography (Devaney, 1984; Williamson, 1991) is a step beyond traveltime tomography, using not just first arrival time data, but phase data from the rest of the scattered wavefield as well. Other techniques involve the use of the wave equation itself to calculate travel times, amplitudes and Frechet derivatives for velocity imaging and inversion (Vidale and Ammon, 1989; Luo and Schuster, 1991). The work of Pratt and Goulty (1991) is a recent example of a popular scheme, which starts with a tomographic analysis to get the smooth

background velocities, then a migration, or imaging step is applied to the full waveform data to get structure. This technique can be applied iteratively to reconstruct both the high (migration-like) and low (smooth background) vertical wavenumber parts of the velocity model (Mora, 1989).

With the sophistication of modern computers, geophysicists now have the capability of using all the information contained in the waveform data for structural imaging and in inversions for the Earth parameters such as velocity and density. Recent work has shown the connection between full waveform inversion and the data processing techniques of migration (Tarantola, 1984; Mora, 1987; Docherty, 1991) and digital filter theory (Cardimona, 1991). Esmersoy and Miller (1989) describe the similarities between the 'backprojection' of a diffraction stack, and the 'backpropagation' in a wavefield extrapolation. Indeed, it is these connections that help us to better understand the mathematics of the inverse problem, and hence help us to gain an intuitive feel for developing innovative solutions. It is certainly important to develop full waveform inversion techniques; however, these techniques must address the limitation inherent in all such geophysical experiments, for example, the insufficient data coverage of the Earth's subsurface. Sometimes we can incorporate into the analysis other geophysical information as constraints to guide the inversion (Lines, et al., 1988). What allows us to infer a realistic model for the Earth from the inadequate data is our a priori knowledge about physically correct Earth models. An important method of specifying the a priori information is through the use of covariance, or smoothing functions (Tarantola, 1984). Squires and Cambois (1991) have recently shown how to describe the model covariance smoothing in a linear tomographic traveltime analysis in terms of filter theory.

The solution to the non-linear inverse problem requires an understanding of the forward theory. Band-limited seismic waves are sensitive to more than just the geometric ray path between the source and the receiver, although most of the contribution to the amplitude in the seismic signal comes from the subset of paths which are close to the classical ray path (the stationary phase path). This result is familiar to seismologists from steepest-descent (asymptotic ray theory) and Radon transform (WKBJ) synthesis of seismograms. It is clear that we must work with more than just the classical ray when describing the full waveform information.

Linearizing the perturbation problem in the forward theory describes how changes in the model affect changes in the data. It is this relationship that is necessary in solving the linearized inverse problem. McGillivray and Oldenburg (1990) give a recent summary of methods for calculating these relationships. By studying the forward problem for full waveform propagation, we can introduce the idea of smoothing functions, and help to understand how to define them for a specific inverse problem.

TECHNIQUES

The forward problem and local linearization

In this section, I will use the notation and procedure of Cardimona (1991), who looked at the forward and inverse problems for a simple model. The forward problem theory gives us a relationship $\mathbf{u}=g(\mathbf{m})$ between the data and the model. I can write this as a Taylor series around some starting model point $\mathbf{m}_o$, given by

$$\mathbf{u} = \mathbf{u}_o + \mathbf{G}\,(\mathbf{m} - \mathbf{m}_o) + O(\delta\mathbf{m}^2)\,. \qquad (1)$$

If I assume that the changes in the model are small enough, I can drop the higher order terms in $\delta\mathbf{m} = (\mathbf{m}-\mathbf{m}_o)$. The result is a local linearization of the function $g(\mathbf{m})$ about the point $\mathbf{m}_o$ given by the operator $\mathbf{G}$. This is the Born approximation, and $\mathbf{G}$ is the linear operator, or Frechet derivative, which maps perturbations in model space ($\delta\mathbf{m}$) to perturbations in the data space ($\delta\mathbf{u}$). Thus we have

$$\delta\mathbf{u} = \mathbf{G}\,\delta\mathbf{m}\,. \qquad (2)$$

I can write the linearized forward problem of equation 2 as

$$\delta\mathbf{u}(\mathbf{r}_o,\mathbf{r}_j,t) = \int_V dV(\mathbf{r}')\,\mathbf{D}(\mathbf{r}_o,\mathbf{r}_j,t;\mathbf{r}')\,\delta\mathbf{m}(\mathbf{r}')\,, \qquad (3)$$

where $\mathbf{r}_o$ is the source position, $\mathbf{r}_j$ is the receiver position and $\mathbf{D}(\mathbf{r}_o,\mathbf{r}_j,t;\mathbf{r}')$ is the Frechet derivative operator kernel. **D** is also known as the 'sensitivity function' in the mathematical literature (Parker, 1977), which makes clear its role in the scattering problem. The Frechet kernel defines the weight given to the model perturbations, and thus describes the sensitivity of the data to changes in the model.

The straightforward way to get the Frechet derivative kernel is to turn the inhomogeneous differential equation for data **u** into an integral equation, using the Green's function solution to the homogeneous equation (e.g., Mathews and Walker, 1970; Appendix I). The solution can be written as a Neumann series, and the first iteration of the series gives the Born approximation, or linearized result. Specifically, the equations of motion for the acoustic scattering problem are given by a differential operator L, such that $L(u) = \delta m\, u$. I define $u=u_i+u_s$, where u_i is the known background field and u_s is the scattered field due to perturbations, δm, in the single model parameter. Changing this to an integral equation for u_s gives

$$u_s(\mathbf{r}_j) = \int G_i(\mathbf{r}_j|\mathbf{r}')\, \delta m(\mathbf{r}')\, u(\mathbf{r}') \,. \qquad (4)$$

Here G_i is the Green's function for the known background medium, i.e. it solves the equation $L(u)=0$. In this equation, $\delta m(\mathbf{r}')u(\mathbf{r}')$ is the secondary force system due to model perturbations at point $\mathbf{r}'$. This type of result is very general, and is routinely used in scattering theory. For example, recently, Kennett (1984) has used it to generalize the reflectivity method in terms of a reflection operator acting as a secondary source function. Assuming that $u_s << u_i$, I can linearize the integral equation above, i.e., apply the Born approximation which neglects higher order terms in u_s (in δm). The result is

$$u_s(\mathbf{r}_j) = \int G_i(\mathbf{r}_j|\mathbf{r}')\, \delta m(\mathbf{r}')\, G_i(\mathbf{r}'|\mathbf{r}_o) \,. \qquad (5)$$

So, the kernel of the Frechet derivative operator $\mathbf{D}(\mathbf{r}_o,\mathbf{r}_j,t;\mathbf{r}')$ in this case is given by $G_i(\mathbf{r}_j|\,\mathbf{r}')G_i(\mathbf{r}'|\,\mathbf{r}_o)$, where as before, $\mathbf{r}_o$ is the source location and $\mathbf{r}_j$ is the

receiver location. The ideas of the Feynman path integral approach (Feynman and Hibbs, 1965) can help give an intuitive feel for the scattering problem. For perturbation problems such as the one just described above, the Feynman path integral approach calculates a summation over all possible scattering paths between the source, the scattering field and the receiver. Linearization of the Feynman path integral result, which is completely analogous to considering only single scattering, gives exactly the same result as in equation 5 where we linearized the Neumann series solution to the integral equation above. As we would expect, linearizing the scattering problem implies that all multiple scattering is neglected.

To illustrate these ideas with an example, a diffraction formula may look like, for a given frequency (Elmore and Heald, 1969),

$$u_s = \iint_S ds \frac{e^{ik_j r_j}}{r_j} \left\{ f(\varphi_j,\varphi_o) \frac{e^{ik_o r_o}}{r_o} \right\}. \qquad (6)$$

Here, the secondary Huygens sources are due to the function $f(\varphi_j,\varphi_o)$ on the surface S. This is the Huygens-Fresnel construction as used in Kirchhoff diffraction theory. We may make an approximation by limiting the surface integral to some Δs, for example about the stationary phase points given by ray theory. We may also make an approximation akin to the paraxial approximation, which would take the angle dependence in the function f outside of the integral, that part being defined by the ray theory stationary point, and leaving only an amplitude factor corresponding to a transmission (or reflection) coefficient.

As seen in the previous example, we may introduce different kinds of appropriate approximations into the forward theory. Given a particular source-receiver distribution and model parameterization, we can define a general smoothing operator **C** as a modification to the true Frechet derivative operator which may include approximations to the forward problem and any a priori information as well. For a source at $\mathbf{r}_o$,

$$\delta\hat{\mathbf{u}}(\mathbf{r}_j,t) = \int_V dV(\mathbf{r}'') \left[\mathbf{D}(\mathbf{r}_o,\mathbf{r}_j,t,\mathbf{r}') \, \mathbf{C}(\mathbf{r}_o,\mathbf{r}_j,t,\mathbf{r}'|\mathbf{r}'') \right] \delta\mathbf{m}(\mathbf{r}'') \; . \qquad (7)$$

The 'product' $\mathbf{D}(\mathbf{r}_o,\mathbf{r}_j,t;\mathbf{r}')\,\mathbf{C}(\mathbf{r}_o,\mathbf{r}_j,t;\mathbf{r}'|\mathbf{r}'')$ may imply an integral or windowing operation. If the operator $\mathbf{C}$ is equal to unity, the result is exactly equation 3 for the original $\delta\mathbf{u}$ data residual. If we choose to introduce smoothing, however, $\mathbf{C}$ becomes non-trivial.

Inverse theory

Using a calculus-based approach to the inverse problem (Appendix III), an iterative Newton method solution, for iteration (n+1), is given by

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \mathbf{H}_n^{-1}\gamma_n \; . \qquad (8)$$

At model point $\mathbf{m}_n$, γ is the direction of steepest ascent, or gradient, and $\mathbf{H}_n$ is the model space curvature operator, or Hessian. In order to minimize a weighted combination of prediction error and model covariance size, the Newton method solution becomes (Tarantola, 1987)

$$\delta\mathbf{m}_{n+1} = \left(\mathbf{C}_m\mathbf{G}_n^t\mathbf{C}_u^{-1} + \mathbf{I}\right)^{-1} \left\{ \mathbf{C}_m\mathbf{G}_n^t\mathbf{C}_u^{-1}\delta\mathbf{u}_n - (\mathbf{m}_n - \mathbf{m}_{n-1}) \right\} \; , \qquad (9)$$

where $\mathbf{G}^t$ performs the transpose operation to the forward problem. Here, $\mathbf{C}_u$ and $\mathbf{C}_m$ are the least squares functional weighting operators, equivalent to the data covariance and model covariance matrices for the discrete problem (Tarantola, 1986). Equation 9 can be simplified to describe the solution for various special cases (Wiggins, 1972; Lines and Treitel, 1984; Menke, 1984) from the generalized inverse for the overdetermined problem, when the covariance operators equal the identity, to the weighted damped least squares solution, when both covariance operators are proportional to the identity.

The kernel for the transpose operator $\mathbf{G}^t$ in the inverse problem is the same as that of the forward problem, except with transposed variables. Using equation 3, we get a model perturbation estimate of the form

$$\delta m_{est}(\mathbf{r}') = \sum_j \int_o^T dt\, \mathbf{D}(\mathbf{r}_o,\mathbf{r}_j,t,\mathbf{r}')\, \delta \mathbf{u}(\mathbf{r}_o,\mathbf{r}_j,t) \; . \tag{10}$$

Given more than one source, the summation of equation 10 would then be over all sources as well.

The operation $\mathbf{G}^t$ may be interpreted in terms of the propagation, or the time delaying of a source estimate relative to the data, and an imaging condition, normally a cross-correlation. Point by point in $\mathbf{r}' = (x',y')$ image space, a source function is forward propagated from the source location $\mathbf{r}_o$ to the image point, the data misfit is back propagated from the receiver position $\mathbf{r}_j$ to the image point, and the resulting pair of time series are correlated to yield an image density. Numerically, the image density is proportional to the estimated changes required in the model parameter. This operation is achieved through the action of the transpose of the sensitivity function upon the data residuals. After Tarantola (1987) I define the *adjoint* $\mathbf{G}^*$ operator to the forward operator $\mathbf{G}$ by

$$\mathbf{G}^* = \mathbf{C}_m \mathbf{G}^t \mathbf{C}_u^{-1} \; , \tag{11}$$

which is the transpose operator, modified by the data and model covariance functions. The application of $\mathbf{C}_u^{-1}$ weights the data directly, while the application of $\mathbf{C}_m$ is a model space smoothing operator. I will concentrate on the adjoint operator since, in the Newton algorithm of equation 9, it is the adjoint which is mapping the data into the model space and smoothing enters into the inverse problem solution through the adjoint operator.

Knowing the exact Born scattering forward problem solution $\delta u = \mathbf{G}_b \delta m$, we then produce the gradient field for the model parameter update with $\mathbf{G}^* \delta u$, or $\mathbf{G}^* \mathbf{G}_b \delta m$. In an imaging experiment, this will be a scaled version of the one-step

gradient solution for δm_{est}. If $\mathbf{G}^*$ is from the exact Born kernel, then $\mathbf{G}^*=\mathbf{G}_b^t$, we have the generalized inverse solution, and $\delta m_{est} = \delta m$ in the limit of infinite bandwidth data and perfect ray path coverage. With finite bandwidth data and more geophysically reasonable ray coverage, we get a band-limited version of the model. When $\mathbf{G}^*$ includes a priori information, we damp the inversion and get a smoothed version of the model corresponding to the application of $\mathbf{G}^*=\mathbf{C}_m\mathbf{G}_b^t\mathbf{C}_u^{-1}$. Equivalently, if we introduce a non-trivial smoothing operator $\mathbf{C}(\mathbf{r}_o,\mathbf{r}_j,t,\mathbf{r}')$, incorporating the data or model covariance information, into the forward theory as in equation 7, we define a new sensitivity function which introduces damping into the inverse problem through the action of the transpose operator.

As a function of $\mathbf{r}_o$, $\mathbf{r}_j$, $\mathbf{r}'$ and t, the operator $\mathbf{C}$ can be designed to achieve a number of smoothing procedures, which can be seen as preconditioning operations in a gradient solution to the inverse problem. In order to suppress the influence of noisy data on the model parameter recovery, we may make the smoothing operator $\mathbf{C}$ be a function of receiver/source positions (Thorson, 1984) or it may be an elaborate residual weighting function (Crase, et al., 1990). Many recent studies have attacked the inverse problem by staging the iterative inversion sequence over such parameters as frequency (Kolb and Canadas,1986), offset (McAulay, 1986) and time. This procedure is simply an application of a different $\mathbf{C}$ operator at each step of the inversion. The staging operations of $\mathbf{C}$, and as a part of the adjoint operation, can be considered part of the linearization kernel, giving a modified transpose operation on the original data set. This modified transpose operation implies a modified (smoothed) forward theory. This is what we might expect, and is exactly what we want to achieve in staging the inversion. For example, we stage in frequency in order to model the low frequencies in the data to invert for the long wavelength variations in the model first, then we move on to invert for the higher frequencies in the model. In another example, Williamson (1990) describes a staged inversion defined by a changing metric related to varying the slowness cell size, and hence the resolution, from iteration to iteration. Also, as in the previous example from Kirchhoff theory, the operator $\mathbf{C}$ can be defined through our understanding of the physical limitations of the forward theory we are using.

As demonstrated in the above examples, we can study the smoothing of the least squares inverse problem in terms of the choice of local linearization $\mathbf{G}$, which implies a transpose operator $\mathbf{G}^t$. We incorporate data and model covariance ($\mathbf{C}_u$ and $\mathbf{C}_m$) information into this transpose operator, defining the adjoint operator to the forward problem. The smoothing and damping that goes into the inverse problem is determined by the choice of adjoint operator. Thus we can look at the single problem of understanding what goes into the forward theory in an effort to understand the inverse problem.

For the tomography problem in the next sections, I consider sensitivity functions in ray-centered coordinates, and I define functions which apply smoothing transverse to the ray. Other researchers have concentrated on the rays in tomography, such as Singh and Singh (1991) who looked at applying constraints to each ray in the tomographic analysis. Tinti and Ugolini (1990) try to get rid of artifacts in their tomography results by a simple process of preselecting appropriate seismic rays before the inversion. Carrion (1991) uses Lagrangian multipliers to constrain the tomographic inversion, improving resolution by decreasing the blurring effects resulting from incomplete angular aperture coverage. My analysis presents a mathematical treatment for applying ray-based smoothing constraints to waveform data, and I illustrate the ideas with an example of smoothing out problems in tomography due to incomplete data coverage. These results could perhaps be coupled with a ray-based model parameterization (Michelena and Harris, 1991) as well, for more efficient calculations.

APPLICATION EXAMPLE

Forward scattering problem

In this section, I describe a simple transmission tomography problem in terms of full waveform inverse theory. I choose to find solutions to the paraxial equation for perturbations from an initially homogeneous model. The parabolic equation is central to the development of the Gaussian beam method, and its applicability requires the use of relatively high frequencies and rays that diverge relatively slowly. Recently, Graves and Clayton (1990) work out frequency domain finite difference solutions for extrapolators in the paraxial wave equation,

and Wapenaar (1990) presented equations for representing seismic sources for use with one-way wave equations. For clarity, I discuss some of the important aspects of this forward transmission problem. Starting from the scalar Helmholtz equation

$$\phi_{xx} + \phi_{yy} + k^2\phi = 0 , \tag{12}$$

I let $\phi(x,y) = \exp(ik_o x)\, P(x,y)$, and introduce the model parameter slowness s such that $\omega s = k$. With a background slowness s_o, I let $(s-s_o)=\delta s$. Putting ϕ into the Helmholtz equation, I get a new differential equation for P(x,y) in which I neglect the P_{xx} term. Noting that $(k^2-k_o^2) = \omega^2(2s_o\delta s + \delta s^2)$ and neglecting the δs^2 term, I get the paraxial equation for P(x,y) given by

$$P_{yy} + i2\omega s_o P_x = -2\omega^2 s_o \delta s\, P . \tag{13}$$

This is a one-way wave equation for functions P(x,y), and it compares directly with the Schrodinger Equation of quantum mechanics. When $\delta s=0$ (free particle), we can get the exact Green's function solution (Mathews and Walker, 1970) and use this to solve the inhomogeneous equation above. After linearizing the problem, I write a solution of the form (for a given frequency ω)

$$\frac{P}{P_o} - 1 = \frac{P_1}{P_o} = \int dx' \int dy'\; K(x,y,\omega,x',y') \left(i\,\omega\,\delta s(x',y') \right) , \tag{14}$$

where $i\,\omega P_o K(x,y,\omega,x',y')$ is the Frechet derivative operator kernel for this problem, acting on slowness perturbations $\delta s(x',y')$ to give perturbations $P_1(x,y,\omega)$ in the data. I note that equation 14 for P_1/P_o is set up directly for extension to the Rytov approximation. If we say that $P=P_o+P_s$, where P_s is the scattered wavefield (P_1), $P_s=P_b$ gives the original linearization implied by the Born approximation. If we say that $P=\exp(\psi_o+\psi_s)$, then we get the Rytov approximation, where $P_o=\exp(\psi_o)$ and $P_b=\psi_s P_o$ (Devaney,1984, Rajan and Frisk,

1989). Hence, in the Rytov approximation, $\psi_s = P_b/P_o$, and $P_b = P_o \ln(P_s/ P_o+1)$. This paper is not one which compares the relative merits of the Born versus the Rytov approximations. Since the Rytov approximation can be achieved through an extension of the Born theory, I restrict this paper to studying the sensitivity function K and the forward problem. For a given frequency, $K(x,y,\omega,x',y')$ of equation 14 is the weight applied to the model perturbation at point (x',y'), as felt at the receiving point (x,y).

Cardimona and Garmany (1989) have worked out the sensitivity function of equation 14. Without loss in generality, I translate the (x,y) coordinates so that the source is at (0,0), and I rotate the coordinates so that the x-axis is the linear ray path between the source and receiver, with X as the distance from the source to the receiver. After making this change of variables, I get a sensitivity function for the Born approximation solution in the form

$$K_b(X,\omega,x',y') = \frac{1}{H}\sqrt{\frac{1}{2\pi}}\exp\left\{ i\left(\frac{(y')^2}{2H^2} - \frac{\pi}{4} \right) \right\} . \qquad (15)$$

H has the units of length and is given by

$$H = \sqrt{\frac{(X-x')\ x'}{X\omega s_o}} . \qquad (16)$$

When $\delta s=0$ in equation 14, we have the free space solution $P = P_o$, and the scattered wavefield $P_1 = 0$, as we expect. When the perturbation in the background slowness is a constant δs_o, then the integral of equation 14 gives, after making the change of variables described above and using equations 15 and 16 with $K = K_b$,

$$\frac{P_1}{P_o} = \int_0^X dx' \int_{-\infty}^{\infty} dy'\ K(X,\omega,x',y') \left[i\ \omega\ \delta s_o \right] = i\ \omega\ \delta s_o\ X . \qquad (17)$$

This is the correct phase change due to the perturbed slowness field, as we see from

$$\frac{P_1}{P_o} = \left(\frac{P}{P_o} - 1\right) = \left\{ \exp(i\,\omega\delta s_o X) - 1 \right\} = i\,\omega\delta s_o X + O(|\delta s_o|^2) \ . \qquad (18)$$

Exchanging the constant slowness perturbation in equation 17 for a linear trend in y', we get no perturbation in the phase or amplitude of the data P, since the y'-integral of the odd linear function against the even Gaussian y' dependence in the kernel K is zero. This is correct for X small compared to the radius of curvature of the perturbed ray. If I choose a quadratic in y' as a slowness perturbation (the harmonic oscillator problem) given by

$$\delta s(x',y') = \frac{a\,(y')^2}{2} \ , \qquad (19)$$

then equation 17 becomes, with $K=K_b$,

$$\frac{P_1}{P_o} = \int_o^X dx' \int_{-\infty}^{\infty} dy'\ K(X,\omega,x',y') \left[i\,\omega \frac{a(y')^2}{2} \right] = - \frac{aX^2}{12 s_o} \ . \qquad (20)$$

We see (from equation 14) that

$$\frac{P}{P_o} = 1 - \frac{aX^2}{12 s_o} \ . \qquad (21)$$

When the second derivative of the slowness is negative ($a<0$ in equation 19, as in a low velocity acoustic channel), equation 21 gives the correct amplitude focusing described by the geometrical spreading equations of ray theory (Cerveny and Hron, 1980). Figure 1 shows the amplitude and wrapped phase for the true Frechet kernel of equation 15. From the phase variation in the transverse coordinate, it is clear the ray 'feels' a region around the classical ray path, and the scale of this region relates to the Fresnel zone size which is dependent upon the wavelength and

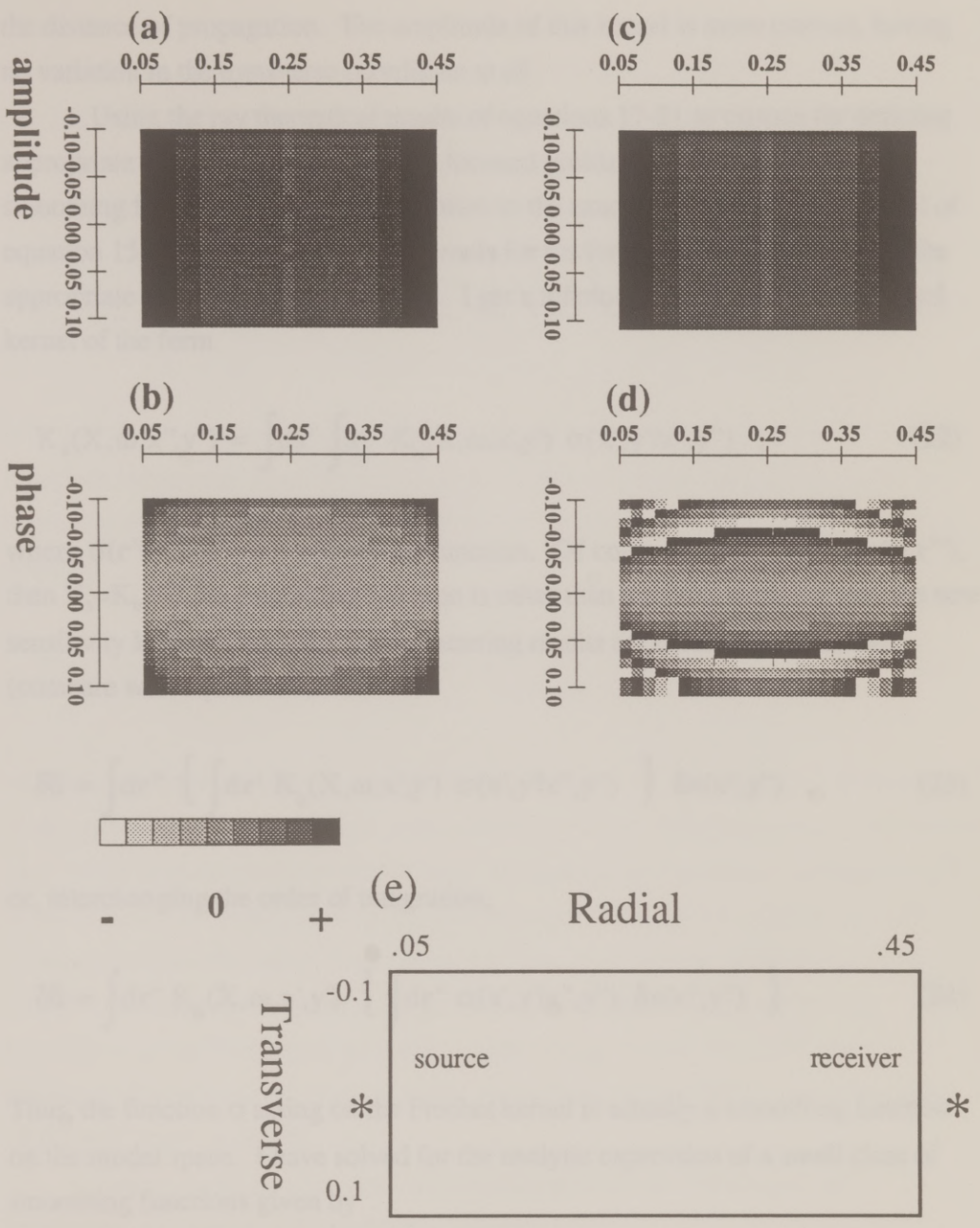


Figure 1) Plots of the amplitude and (wrapped) phase for the Born kernel of the paraxial wave equation at specific frequencies: (a) and (b) 10Hz; (c) and (d) 60Hz. The functions are plotted in ray-centered coordinates. (e) The geometry for a source at (0,0) and a receiver at distance X=0.5 km from the source. All spatial coordinates in km.

the distance of propagation. The amplitude of this kernel is more curious, having no variation in the transverse coordinate at all.

Using the ray theoretical results of equations 17-21 as criteria for defining appropriate sensitivity kernels for the forward problem, I can define some smoothing functions which, when applied to the exact Frechet derivative kernel of equation 15, create new sensitivity kernels for the forward problem that retain the appropriate analytic scattering results. I get a relationship for the new, smoothed kernel of the form

$$K_s(X,\omega,x'',y'') = \int dx' \int dy' \; K_b(X,\omega,x',y') \; \sigma(x',y'|x'',y'') \quad , \tag{22}$$

where $\sigma(\mathbf{r}'|\mathbf{r}'')$ is some smoothing function. Of course, if $\sigma(\mathbf{r}'|\mathbf{r}'') = \delta(\mathbf{r}' - \mathbf{r}'')$, then $K_s=K_b$. If the smoothing function is other than the delta function, and the new sensitivity kernel gives the correct scattering results as described earlier, then (compare with equation 7)

$$\delta\hat{u} = \int d\mathbf{r}'' \left\{ \int d\mathbf{r}' \; K_b(X,\omega,x',y') \; \sigma(x',y'|x'',y'') \right\} \delta s(x'',y'') \quad , \tag{23}$$

or, interchanging the order of integration,

$$\delta\hat{u} = \int d\mathbf{r}' \; K_b(X,\omega,x',y') \left\{ \int d\mathbf{r}'' \; \sigma(x',y'|x'',y'') \; \delta s(x'',y'') \right\} \quad . \tag{24}$$

Thus, the function σ acting on the Frechet kernel is actually a smoothing function on the model space. I have solved for the analytic expression of a small class of smoothing functions given by

$$\sigma(x',y'|x'',y'') = \delta(x'-x'') \sqrt{\frac{1}{2\pi}} \frac{1}{F} \exp\left\{ \frac{y''}{FH} y' - i\frac{\pi}{4} \right\} \quad , \tag{25}$$

where H is the same as in equation 16, and F is another function of X, x', s_o and ω. In equation 15 for the correct sensitivity function K_b, H gives the ray-parallel (x') dependence. The smoothing operator of equation 25, when used as in equation 22 to get a new sensitivity kernel, exchanges H with the function F in equation 15. That is, using equations 25 and 15 to solve for a new kernel K_s, we get

$$K_s(X,\omega,x',y') = \frac{1}{F}\sqrt{\frac{1}{2\pi}}\exp\left\{ i\left(\frac{(y')^2}{2F^2} - \frac{\pi}{4}\right)\right\} . \qquad (26)$$

When F = H in equation 25, the result of equation 26 is K_b of equation 15. I can introduce three other possible functions for F that still retain the correct ray theory scattering results described earlier. These are (to compare with equation 16 for H) given by

$$F = \sqrt{\frac{x'}{3\omega s_o}}, \; F = \sqrt{\frac{(X-x')}{3\omega s_o}} \text{ and } F = \sqrt{\frac{X}{6\omega s_o}}, \qquad (27)$$

which describe, respectively, near-source, near-receiver, or far-field x'-dependence in the sensitivity kernel. Figure 2 shows the amplitude and phase plots at 30 Hz for each of the kernels given by the different choices for the function F in equation 27. These kernels are examples where I have changed the sensitivity function of the forward problem for use in the inverse problem. Application of the near-source kernel for example, because the square-root singularity (and hence the high sensitivity) at the receiver is removed, might help to smooth out problems due to inaccurate knowledge of receiver positions within a well, or inaccurate knowledge of the well position itself.

I have found another class of smoothing functions which give very interesting, ray-theory-like sensitivity kernels:

$$\sigma(x',y'|x'',y'') = \delta(x'-x'')\sqrt{\frac{1}{2}}\,\frac{1}{F}\exp\left\{ -\left(\frac{H\,y''}{F\,y'}\right)^2 - i\frac{\pi}{4}\right\} . \qquad (28)$$

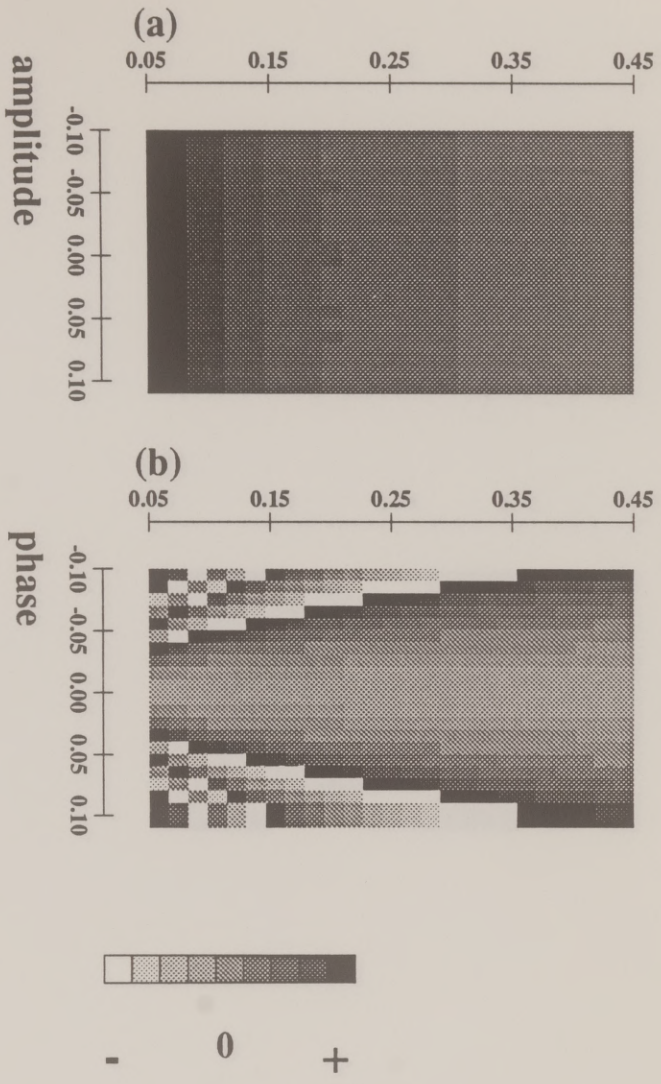


Figure 2) (a) and (b) Plots of the amplitude and (wrapped) phase at 30 Hz for a near-source response sensitivity function associated with the paraxial wave equation. The ray-centered coordinate geometry is as in Figure 1 (e).

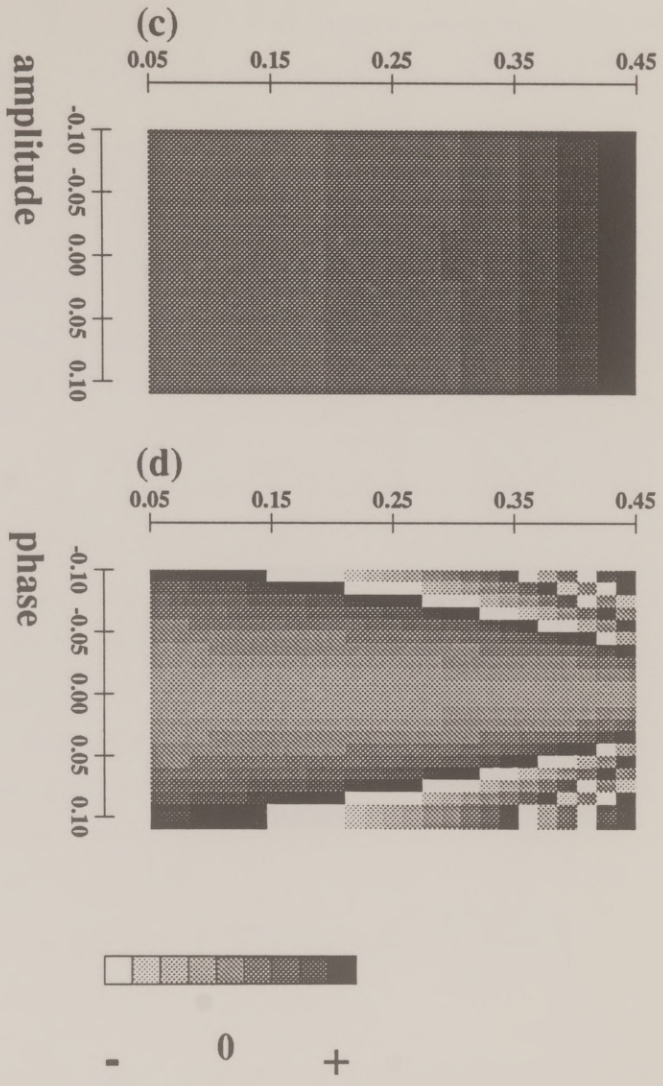


Figure 2) (c) and (d) Plots of the amplitude and (wrapped) phase at 30 Hz for a near-receiver response sensitivity function associated with the paraxial wave equation. The ray-centered coordinate geometry is as in Figure 1 (e).

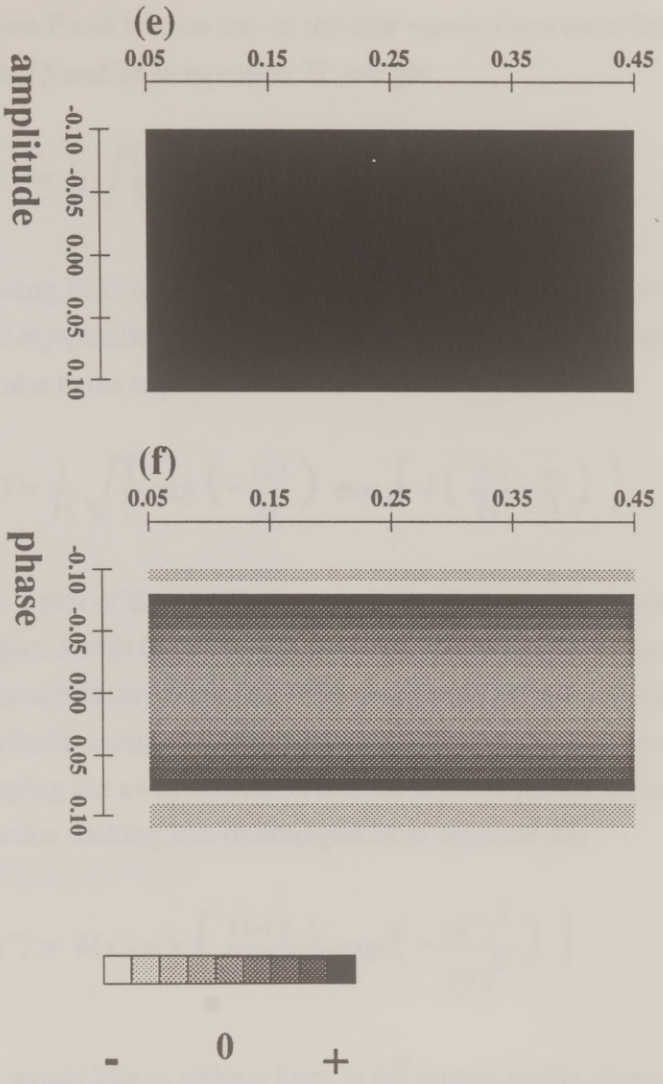


Figure 2) (e) and (f) Plots of the amplitude and (wrapped) phase at 30 Hz for a far-field response sensitivity function associated with the paraxial wave equation. The ray-centered coordinate geometry is as in Figure 1 (e).

Here, the function F can take on any of the four possibilities described earlier. Using equations 15 and 28 in equation 22, we get

$$K_s(X,\omega,x',y') = \frac{1}{F}\sqrt{\frac{1}{2}}\exp\left(-\frac{|y'|}{F}\right)\exp\left\{ i\left(\frac{|y'|}{F}-\frac{\pi}{4}\right)\right\} . \qquad (29)$$

For example, using F=H in equations 28 and 29, we have a sensitivity function with the correct ray-parallel (x') dependence, but including an exponential damping term perpendicular to the ray:

$$\tilde{K}(X,\omega,x',y') = \frac{1}{H}\sqrt{\frac{1}{2}}\exp\left(-\frac{|y'|}{H}\right)\exp\left\{ i\left(\frac{|y'|}{H}-\frac{\pi}{4}\right)\right\} . \qquad (30)$$

Figure 3 shows a plot of the amplitude and phase of the kernel of equation 30. Comparing Figure 3 with Figure 1, it is clear that the main difference is in the amplitude of the sensitivity functions. The smoothed function of equation 30 has a transverse amplitude variation which also scales as the Fresnel zone size, with the sensitivity damping out away from the classical ray. To get equation 30, I used a smoothing function looking like (rearranged from equation 28)

$$\sigma(x',y'|x'',y'') = \delta(x'-x'')\left\{ \frac{(1-i)}{2}\frac{1}{H}\exp\left(-\frac{(y'')^2}{(y')^2}\right)\right\} . \qquad (31)$$

What I would like to address here, is the impact on the inverse problem of the difference in the sensitivity kernels. Using the kernel K-tilda of equation 30, we see that the forward problem is given by scattering defined by the true linearized scattering kernel K_b, but with a smoothed version of the model perturbation given, using equation 31, by

$$\delta\tilde{s}(x',y') = \frac{(1-i)}{2}\frac{1}{H}\int dy''\exp\left(-\frac{(y'')^2}{(y')^2}\right)\delta s(x',y'') . \qquad (32)$$

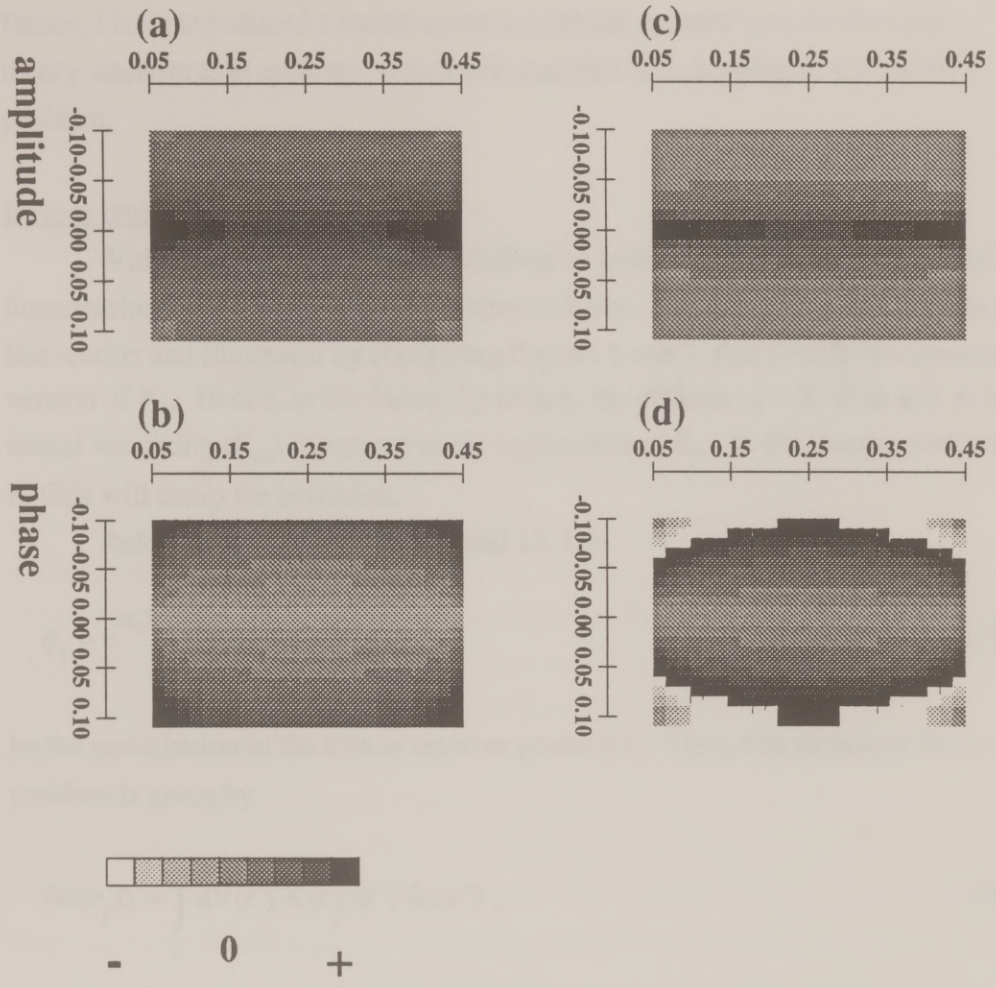


Figure 3) Plots of the amplitude and (wrapped) phase for a smoothed sensitivity function associated with the paraxial wave equation at specific frequencies: (a) and (b) 10 Hz; (c) and (d) 60 Hz. The ray-centered coordinate geometry is as in Figure 1 (e).

Hence, I have introduced a model space smoothing operator into the forward theory, transverse to each ray, which will manifest as a damping in the inverse problem.

Inverse problem

With the sensitivity functions defined as in the last section, I put the local linearization back into the inverse problem solution. It is clear, as pointed out in the last section and illustrated by comparing Figures 1 and 3, that K-tilda is a smoothed version of K_b. Hence, in the forward problem, the difference in K-tilda acts as a model weighting ($\mathbf{C}_m$) compared to the application of K_b. In the inverse problem, K-tilda will damp the inversion.

Refering back to equation 12 and 13, I let

$$\phi_1 = e^{ik_oX} P_1(x,y) = \delta u(\mathbf{r}_j,t) \tag{33}$$

be the perturbation in the data at receiver position $\mathbf{r}_j$. Then, this particular forward problem is given by

$$\delta u(\mathbf{r}_j,t) = \int_V dV(\mathbf{r}') \hat{K}(\mathbf{r}_j,t,\mathbf{r}') \delta s(\mathbf{r}') , \tag{34}$$

where I have calculated the forward response for a set of frequencies, and then applied the Fourier transform to get the time dependent kernel

$$\hat{K}(\mathbf{r}_j,t,\mathbf{r}') = \int d\omega\, e^{i\omega t} \left\{ i\omega W(\omega)\, e^{ik_oX_j} P_o(X_j,0)\; K(X_j,\omega,x',y') \right\}. \tag{35}$$

$W(\omega)$ is the source wavelet signature, and K is one of the sensitivity functions described earlier. The transpose of the Frechet derivative kernel corresponds to the application of the same sensitivity function on the data, integrating over the transposed variables. Thus, the transpose operation looks like

$$\delta s(\mathbf{r}') = \sum_j \int_0^T dt\, \hat{K}(\mathbf{r}_j,t,\mathbf{r}')\, \delta u(\mathbf{r}_j,t)\ . \qquad (36)$$

The summation of equation 36 is over all source-receiver-pair data traces 'j'.

Imaging experiments

I have calculated the first iteration of the gradient of the model update for two synthetic data experiments in a cross-hole waveform tomography problem, comparing the results using the exact sensitivity kernel K_b from equation 15 and a smoothed version K-tilda from equation 30. I show the geometry of the tomographic experiments in Figure 4. I calculate the synthetic seismograms in a 10-60 Hz band, for each of five shots and five receivers using equations 14, 15 and 33. Then I use equation 36 for each of the sensitivity functions, to get an image of the slowness perturbation and to see the effects of each sensitivity kernel. The results for δs = a constant are shown in Figure 5, and results for δs = a point perturbation in the middle of the model are shown in Figures 6 and 7. The top three rows of images in Figures 5 - 7 are the contributions to the gradient from selected single shot gathers. In Figure 7, I show the result after using K_b as the sensitivity function, but with restricting the modeling to a 5-30 Hz band (as in a frequency-staging step). Figures 5 - 7 show that K-tilda acts to damp the inversion, making the imaging results more well behaved.

In Figure 6, the exact result is clearly the best; although it still shows the horizontal smearing of the point perturbation due to lack of surface data. The difference between the results is most dramatic in the single shot gather images, where the data coverage is poor. The full five-shot image results in Figure 7 for the 'long-wavelength' inversion are more like the smoothed results using K-tilda; however, it is clear from the separate shot gathers that using the smoothed K-tilda sensitivity function does a better job of damping the inversion.

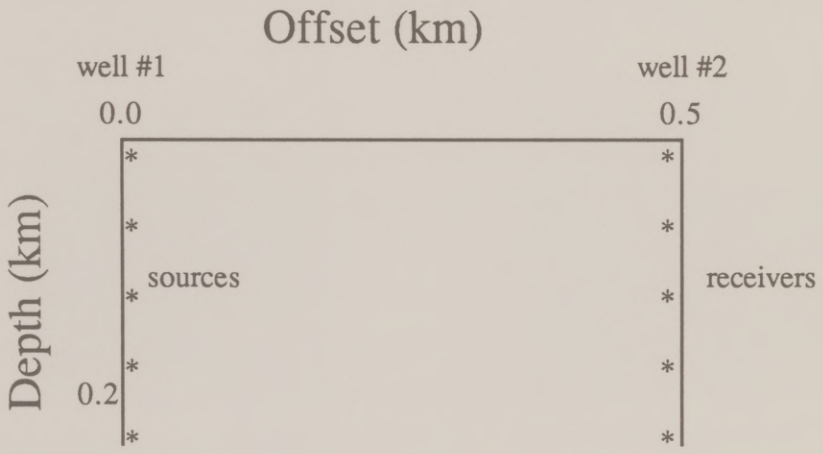


Figure 4) Geometry of a simple cross bore-hole tomographic experiment. Five sources at 50 m spacing in well #1 and five receivers at 50 m spacing in well #2 gives 25 source-receiver pair records.

Figure 5) Example gradient calculation (imaging experiment). The relative grey-scale is as in Figures 1-3. The slowness perturbation for the forward synthetics is taken as a small constant positive value, and the synthetics are calculated for the frequency band 10-60 Hz. (a) - (d): Results using the true Born kernel; (e) - (h): Results using the smoothed sensitivity function K-tilda. For comparison, (a) and (e) are the contributions to the gradient from the record gather for shot #1; (b) and (f) for shot gather #3; (c) and (g) for shot gather #5, for the Born and smoothed kernels respectively. The full gradient update (the image) of perturbations to the background slowness is given by a sum of the contributions from all five shots, and these are shown in (d) and (h). The cross-hole geometry is as in Figure 4. The inversion damping associated with the use of K-tilda in (e) - (f) is most pronounced where data coverage is sparse, as seen in the single shot gather imaging and near each well in the full gradient results.

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
0.05 0.15 0.25 0.35 0.45
0.00 0.05 0.10 0.15 0.20
shot #1
shot #3
shot #5
gradient

Figure 6) Example gradient calculation (imaging experiment). The grey-scale is as in Figures 1-3. The slowness perturbation for the forward synthetics is taken as a point perturbation at (.25,.10). This point in image space is well covered by the source-receiver ray paths. The cross-hole geometry is as in Figure 4, and the shot gather and data description for (a) - (h) is as in Figure 5. The inversion damping associated with the use of K-tilda in (e) - (h) is again most evident where data coverage is sparse, as with the single shot gather imaging. The smoothed results of (e) - (h) are qualitatively more well behaved, and the full imaging of the perturbation in (h) is more localized in the x coordinate than the comparable image in (d) using the true sensitivity kernel.

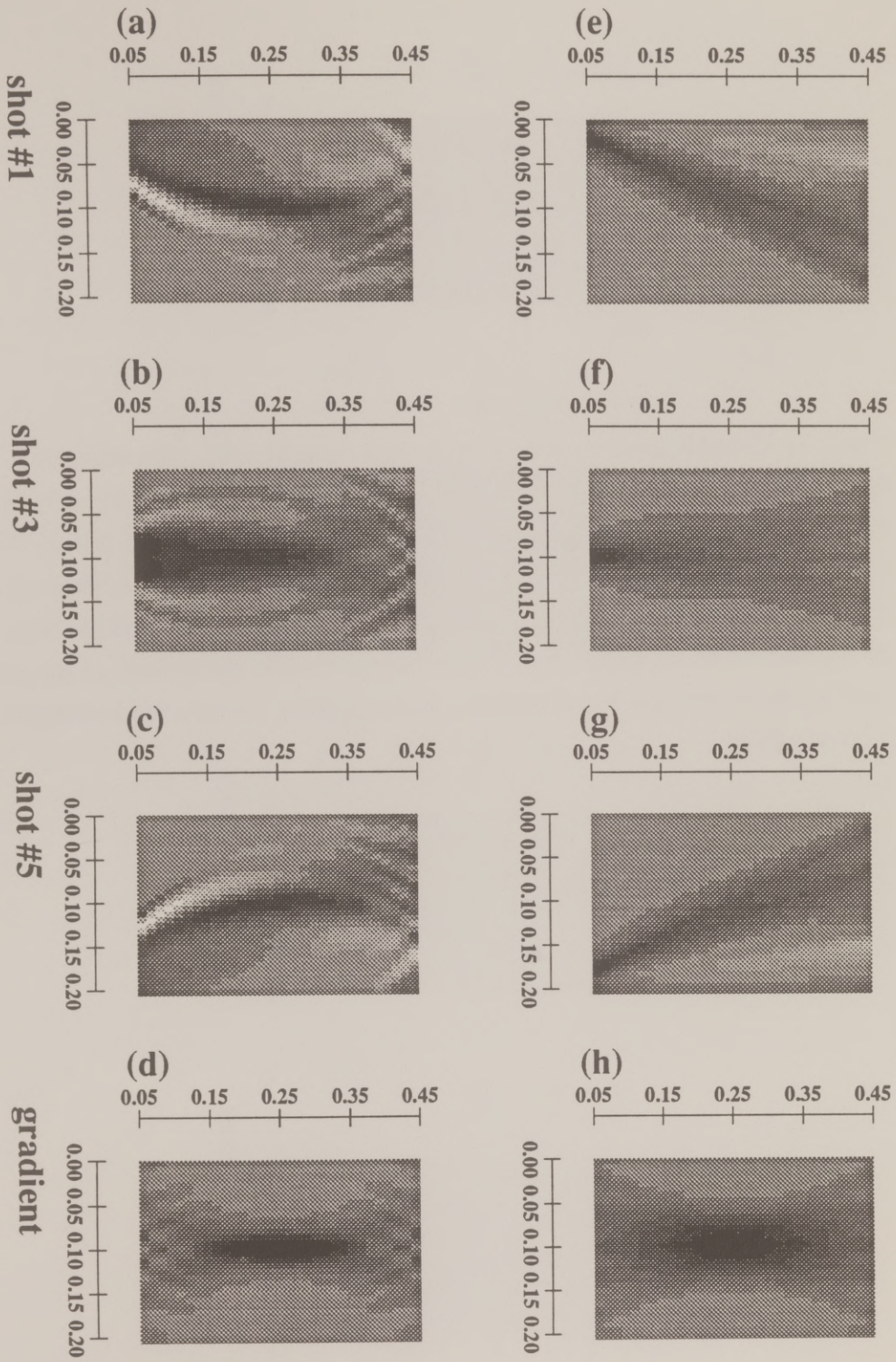
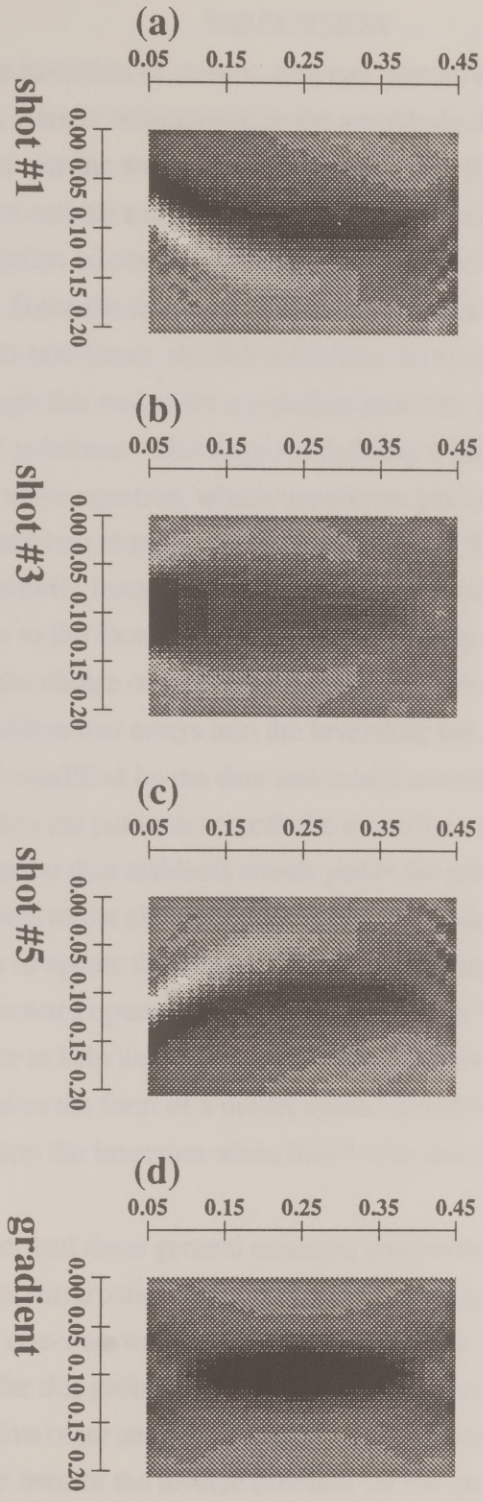

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
0.05 0.15 0.25 0.35 0.45
0.00 0.05 0.10 0.15 0.20
shot #1
shot #3
shot #5
gradient

Figure 7) Example gradient calculation (imaging experiment). The slowness perturbation, experimental geometry and synthetic calculation is exactly as in Figure 6 (a) - (d), except the synthetics are calculated for a frequency band 5-30 Hz. The inversion result of (d) is similar to the smoothed result of Figure 6 (h), although the damping was introduced in a completely different manner. (d) still suffers from poor resolution in the x direction. Where data coverage is sparse there are still the same problems associated with using the true sensitivity function for the paraxial wave equation.


(a)
0.05 0.15 0.25 0.35 0.45
shot #1
0.00 0.05 0.10 0.15 0.20
(b)
0.05 0.15 0.25 0.35 0.45
shot #3
0.00 0.05 0.10 0.15 0.20
(c)
0.05 0.15 0.25 0.35 0.45
shot #5
0.00 0.05 0.10 0.15 0.20
(d)
0.05 0.15 0.25 0.35 0.45
gradient
0.00 0.05 0.10 0.15 0.20

DISCUSSION

Traveltime inversion of seismic data can give an estimate of the gross Earth model. However, there is information in the amplitude data as well as the phase. Also, band-limited seismic waves are sensitive to more than just the ray path between the source and the receiver. Inversion of seismic waveforms gives a solution to the problem of obtaining the best Earth model which explains the full data information. Since the relationship between the data and the Earth parameters in the true model is non-linear, the full waveform inversion of the data must be nonlinear. Although this makes for a complex problem, iterative gradient methods can give 'realistic' solutions. This requires defining a linear operator, based on some form of the wave equation, which transforms perturbations in a particular set of model parameters into changes in the recorded data. This is the Frechet derivative, or sensitivity function. By analyzing the forward theory, and hence the transpose operator to the forward problem, we can study damping in the inverse problem through the choice of local linearization. I have shown that it is the adjoint of the forward problem that enters into the inversion; but, the adjoint is the transpose operator modified by the data and model covariance information. The covariance operators are put in to smooth the inversion. It is the action of the adjoint operator on the data residuals which yields the gradient of the objective function with respect to the chosen model parameters, where the gradient gives the direction in which to update the model. When the adjoint is other than the exact transpose of the forward operator, it acts to precondition the gradient in the inverse calculation in order to help the convergence of the iterative scheme. This preconditioning takes the form of a model space smoothing function, and can be defined to help damp the inversion when faced with specific problems in the data set.

I have described these general concepts and investigated, as an example, the effects of slowness perturbations from an initially homogeneous model on solutions of the paraxial, or one-way wave equation. I have given the exact Frechet derivative kernel for this problem, and I have also defined some smoothing operators which give other sensitivity functions for the forward problem. When these operators are used in the inverse problem for the gradient update, the

smoothing functions act to precondition the gradient, manifesting as smoothing transverse to each ray. After casting this in the form of a tomographic cross-hole seismic experiment, I show the results from two synthetic data cases. With an example gradient calculation for each case, I demonstrate how the smoothed sensitivity functions can be used to help achieve a reasonable imaging result when ray coverage is insufficient.

Knowledge of the forward problem solution and the approximations we introduce into that calculation provide us with an understanding of how the inverse problem will behave. In imaging experiments, the variation in model parameters is sought, and the smoothing introduced via the adjoint operator can help attain a reasonable result. Of course, we have to be more careful in choosing smoothing functions as we proceed through the iterations for an inversion for the exact model parameter values.

The most important criterion for evaluating a scientific theory is the success of predictions based on that theory.

--Stephen Brush (from *Prediction and Theory Evaluation*, EOS, 1990)

CHAPTER 3: SEISMIC SCATTERING IN THE UPPER MANTLE

INTRODUCTION

In a seismological study using deep earthquake records, there are certainly a number of physical problems which might be addressed. The mantle constitutes over 80 percent of the Earth; thus, describing mantle processes is very important for our understanding of the Earth's evolution. In this chapter, I focus on near source scattering and upper mantle structure in the Tonga subduction zone. Upper mantle attenuation in the subduction zone environment is fairly well understood (e.g., Barazangi, et al., 1975), and can help to constrain mantle temperatures; although, researchers might try to do more anelastic studies of the transition zone to help characterize the temperature structure around the 670 km discontinuity. Determining the extent to which the Earth is chemically stratified, and understanding the density structure which controls mantle dynamics, are two of the more pressing issues. Intuition into the nature of upper mantle discontinuities is important for describing upper mantle models.

The long wavelength features of the compressional and shear wave velocities from travel time analyses and wave form modeling, and the density structure from free oscillation data agree fairly well. Surface wave tomography has given us enticing pictures of the upper mantle. Both dispersion analysis through a spherical harmonic expansion of phase and group velocities (Nataf et al., 1986) and full waveform inversion of fundamental and higher mode surface waves (Woodhouse and Dziewonski, 1984) give very similar long wavelength features in models of upper mantle velocity structure. The resolution of seismic imaging techniques is dependent upon the frequency band of the data, the ray path coverage and the length of the ray paths. Tomography with surface waves cannot resolve

structures smaller than around 4000 km because it uses data with periods greater than 100 seconds. Normal mode studies of the free oscillations of the Earth can help to determine the departures from spherical symmetry within the Earth, but again these are very long wavelength features. Global travel-time inversions and waveform modeling of body waves have given velocity structure, yet resolution is still poor since low frequencies and long travel distances are most often used. It is in studying the short wavelength, lateral variations in mantle properties from which we may determine the more dynamic features of the mantle. Seismology has been the leading geophysical technique for determining the smooth features of the Earth, and it is seismology which can help to characterize the departures from the smooth models. We must use specialized data sets which have sampled specific areas of interest.

The upper mantle

The smooth models of the mantle have not changed much over the years, and can mostly be described in terms of a homogeneous material under compression due to its own weight. In the upper mantle, the low-velocity zone (LVZ), related to the temperature structure, and the transition zone between 400 and 800 km, related to changes in mineral phases or compositional changes, are two anomalous regions. The LVZ is important in the study of the basaltic source region and the variability of the zone can have great petrologic significance (Anderson,1987). The large lateral variations in seismic velocities in the upper mantle are most likely due to variations in the depth, thickness and intensity of the LVZ (Bott, 1982). Grand and Helmberger (1984a,b) have shown the variation in upper mantle shear velocities for different types of regions. The features in the region of the mantle between 200-700 km are predominantly controlled by phase changes in the system Mg_2SiO_4 - Fe_2SiO_4 (Anderson, 1967), which give the olivine to spinel to perovskite transitions. There may be other transitions near these depths. Liu (1980) has proposed a phase transformation associated with the low pressure boundary of the two-phase region for the pyroxene to garnet transition which could yield a seismic discontinuity in the vicinity of 200 km. The high pressure boundary for this transition may explain a discontinuity at 520-570 km. If

this is not the cause, then a mechanism such as a chemical change must be invoked to explain any seismic discontinuity at these depths.

From seismic data, the relative significance of the compositional influences is hard to evaluate directly. Partial melting and the trade-off of composition versus temperature effects in transition regions make it difficult to determine the composition of the oceanic upper mantle between 50-400 km (Anderson and Bass, 1984, Anderson, 1987). However, seismic data may help to constrain temperature structure, and it may help researchers to understand the structure of any discontinuities, offering insight into the problem of whether they are chemical or phase boundaries.

Investigators have looked at the PP and the P'P' (PKPPKP) phases and have recognized that precursors to these phases could be seismic reflections from discontinuities below the surface of the Earth (Adams,1968; Bolt et al., 1969; Engdahl and Flinn, 1969; Whitcomb and Anderson, 1970). These seismic studies were able to help identify reflections from numerous boundaries in the mantle, one of the strongest being off of the 670 km discontinuity. Amplitude ratio work of Whitcomb and Anderson (1970) suggests that, for periods of one to two seconds, a first order discontinuity model is a good approximation to upper mantle reflectors, and their observed reflections require a transition zone of no more than 4 km thickness. King et al. (1976) argue against the model of first order discontinuities in the upper mantle, except for the 670 km discontinuity. Their time series analyses, and other similar work (e.g. Haddon et al., 1977), has shown that random scattering by inhomogeneities in the crust and upper mantle could explain precursors to PP and P'P', while scattering at the core-mantle boundary can account for the precursors to PKIKP.

Ward (1978) calculated long-period synthetic seismograms of reflections and conversions due to discontinuities at 420 and 671 km to compare with data from a New Hebrides shallow earthquake. He explains how coherency in the data record section argues against a random scattering origin for all of the 'coda phases'. His results indicate that reflections and conversions due to upper mantle discontinuities can be a prominent part of the seismic record in the time window

between P and PP. Davis et al. (1989) have used broad-band data to study the P'P' precursor idea, finding only one candidate P'670P' phase.

Other analysts have used a variety of seismic data to add evidence in support of upper mantle structure. Hales et al. (1980), fit refraction curves to short period array data in Australia to constrain the depths of upper mantle velocity discontinuities. More recently, Revenaugh and Jordan (1987,1989) used ScS reverberation phases in a long period data modeling and inversion analysis from which they associated peaks in a one-dimensional reflectivity series with upper (or lower) mantle structure variation. Shearer (1990a) stacked a huge amount of long period digital data from shallow earthquakes around the world into record sections, showing that phases from the 410 and 670 km discontinuities were indeed visible and globally coherent. As with many previous analyses, long-period, globally smoothed seismic analyses cannot show any specific deviations from an average Earth model.

Since our knowledge about the deep interior of our planet has mainly come from indirect evidence, it is not completely surprising that the old question of whether or not the mantle is chemically layered has yet to be decided. Temperature studies argue in favor of the two-layer model. In a model with two distinct convecting layers, conduction would be the mechanism for the transfer of heat between the lower and upper mantle. That could perhaps explain the retention of heat within the Earth, since conduction is less efficient than convection. Studying volcanic rocks, geochemists deduce from isotopic analyses that there are reservoirs of distinctly different rocks in the mantle. This is easily described by a two-layer model of a depleted upper mantle within which are occasional plumes of primitive lower mantle material. The geochemical 'plum-pudding' model, which exhibits varying sized reservoirs of different compositions scattered throughout the mantle, could still fit in with a model of whole mantle convection.

Seismology has given new evidence to add strength to arguments for whole mantle convection. Although earthquakes abruptly stop at depths of around 680 km (Stark and Frohlich, 1985), recent studies of seismic traveltime anomalies supports aseismic extensions of the descending slabs down to at least 1000km (Creager and Jordan, 1986). Silver and Chan (1985) assume that the complex seismic signals

from deep earthquakes are due to the focussing of energy by a slab penetrating well below the 670 km depth.

Tantalizing as these seismic studies seem, seismology has also offered evidence against whole mantle convection. Giardini and Woodhouse (1984) have noted substantial slab deformation at or near the 670 km discontinuity, implying at least quite a bit of resistance of the subducting slab to penetration into the lower mantle. Convection models that try and reconcile gravity measurements with the seismicity and stress orientation data tend to prefer whole mantle mixing, although it does seem that the lower mantle must be substantially more viscous than the upper mantle (Hager, 1984). This would explain why the slab might meet with some resistance in penetrating below 670 km depths. Not only are there arguments which use mineralogy and anisotropy of the slab minerals to account for the seismic traveltime anomalies associated with subduction zone earthquakes (Anderson, 1989), but Grand (1989) has shown that errors in the traveltime data used with the residual sphere analyses can introduce artificial azimuthal anomalies unassociated with slab penetration into the lower mantle. Weber (1990) has shown recently that studying the amplitude variations in seismic signals coming from subduction zone earthquakes may be very useful in determining lateral heterogeneity and slab structure.

With data from subduction zone earthquakes, seismologists have identified and used the S to P converted phase at the 670 km discontinuity to help characterize this region of the transition zone in the direct vicinity of subduction (Barley et al., 1982; Richards and Wicks, 1990, Bock and Kind, 1991). It seems from these analyses that the 670 km discontinuity does not have a lot of topography within the vicinity of the slab (Richards and Wicks, 1990; Wicks and Richards, 1991). This adds evidence against introducing a chemical change to describe the discontinuity, since that would imply a very broad downwarping of the boundary. The slab would create quite a disturbance as it met resistance to penetration into the lower mantle.

It is clear that the mantle must be chemically heterogeneous; but it is hard to determine on what scale this is true. From the analysis of earthquake seismic data, it seems we cannot resolve whether the slab penetrates into the lower mantle or not;

hence, we cannot determine whether a stratified mantle model is more appropriate than models with whole mantle convection. Recent convection calculations which include a phase change at 670 km show whole mantle convection disrupted by the phase transition (Machetel and Weber, 1991). Thus, describing the 670 km discontinuity solely in terms of an endothermic phase change from spinel to perovskite might reconcile the seismic observations, giving a model of whole mantle convection that is disrupted due to the mineral phase change and only allows intermittent mixing between the upper and lower mantle.

An upper mantle data set

From the records of the International Seismological Centre (ISC), there are over 100 occurrences of deep (depth>200 km) earthquake events with body-wave magnitude greater than or equal to 5.5 between January 1980 and December 1985. About half of these events occur in the Tonga subduction zone near Fiji Island. I have examined short period (around 2 sec) vertical component seismic records of the North American stations of the Global Digital Seismograph Network (GDSN). My data set consists of 129 records from 46 events (Table 1). I used the ISC locations and origin times, with National Earthquake Information Center parameters for the events in 1986. Figure 8 shows the data records aligned at the P wave and plotted against source depth, with amplitudes equalized between the different sources. The original data records have a sample rate of .05 sec, so the Nyquist frequency is 10 Hz. In Figure 9, I show the pP reflection points on the surface of the Earth to show the data record coverage, plotted with the bathymetry of the region (van Wykhouse, 1973).

I concentrate on the time window between the P and pP phases. Because this time window is uncluttered by other phase arrivals, I can interpret coherent energy in the P coda as due to scattering from upper mantle discontinuities. The pP phase represents the free surface reflection above the earthquake, and I use this information to get a reliable source depth estimate. After associating the P phase with the source depth, I can map energy after the first arrival to scattering from boundaries above and below the earthquake depth . This defines three separate analyses: I can interpret the coda energy as P-P reflections from discontinuities

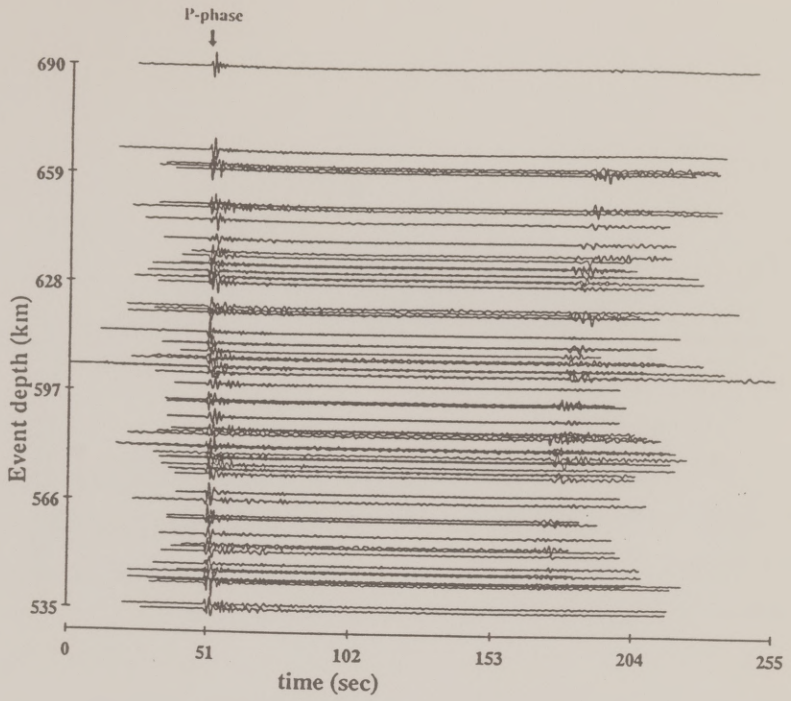
Table 1

Date	Time	Lat.	Long.	mb	d	h	stations used
Jan. 15, 1980	04:25:44.9	22.17S	179.5W	5.5	594 I	613.9	A
Apr. 22, 1980	06:55:28.9	18.24S	178.09W	5.5	589 I	604.4	A
Jun. 14, 1980	19:32:35.8	18.28S	177.95W	5.5	549 I	576.5	A
Jun. 17, 1980	08:42:56.9	20.21S	178.44W	5.5	580 I	624.4	A
Jul. 20, 1980	21:20:03.6	17.89S	178.60W	6.0	588 I	605.8	A
Dec. 26, 1980	05:51:44.9	22.18S	179.56W	5.5	583 I	596.4	A
Feb. 6, 1981	07:23:07.0	21.11S	178.94W	5.6	624 I	641.6	A
Apr. 25, 1981	05:36:41.3	22.23S	179.42E	5.6	586 I	632.6	A
Apr. 28, 1981	21:14:47.1	23.70S	179.99E	6.0	522 I	544.4	L
Jul. 07, 1981	23:24:49.0	22.91S	179.46W	5.5	535 I	546.2	A
Aug. 17, 1981	17:07:40.9	25.40S	179.07W	5.5	376 I	399.5	A, L
Oct. 07, 1981	03:02:12.7	20.70S	178.65W	5.9	606 I	630.9	A
Oct 20, 1981	13:51:45.7	18.25S	179.30W	5.5	641 I	656.8	A, L
Nov. 25, 1981	23:51:15.4	20.81S	178.87W	5.8	594 I	625.0	A, L
Apr. 19, 1982	07:55:15.8	26.07S	178.62W	5.5	344 I	353.8	A, L
May 12, 1982	10:03:31.2	24.63S	179.22E	5.6	536 I	569.4	A, L
Jul. 05, 1982	21:22:26.7	20.78S	178.80W	5.5	609 I	626.5	A, C, L
Sep. 17, 1982	13:28:24.8	23.45S	179.87W	5.9	543 I	571.4	A, C, L
May 02, 1983	09:58:13.6	20.65S	178.49W	5.6	594 I	608.2	C, L
Sep. 16, 1983	08:09:27.1	23.98S	179.80W	6.0	515 I	535.4	A, C, L
Nov. 06, 1983	09:38:40.1	20.12S	177.69W	5.5	388 I	399.8	A, C, SD
Nov. 29, 1983	23:41:07.5	19.45S	177.79W	5.6	526 I	536.0	A, C, L
Jan. 19, 1984	16:15:15.6	23.68S	178.30W	5.8	326 I	339.9	A, C, L, SD
Mar. 12, 1984	10:50:48.9	23.41S	179.99W	5.5	546 I	574.5	A, C, L
Mar. 14, 1984	11:36:30.8	20.10S	178.07W	5.7	567 I	588.4	A, C, L
Apr. 22, 1984	03:33:00.5	21.83S	179.37W	5.7	586 I	628.7	A, C, L
Apr. 25, 1984	04:19:30.8	17.31S	177.20W	5.6	404 I	421.8	A, C, L, SD
Jun. 15, 1984	14:22:25.4	15.83S	174.80W	6.0	267 I	265.5	A, NT
Jul. 03, 1984	13:42:00.3	17.69S	178.87W	5.5	526 I	540.6	A, C, L
Aug. 26, 1984	05:00:45.8	23.56S	179.00E	5.7	554 I	593.0	A, C, J, L
Sep. 28, 1984	03:03:46.7	21.43S	177.79W	5.7	357 I	378.6	A, C, J, L, SD
Oct. 10, 1984	19:05:58.2	20.11S	179.28W	5.6	676 I	689.6	A, J
Nov. 15, 1984	05:52:30.3	20.37S	177.41W	5.6	346 I	358.2	A, C, L, SD
Nov. 17, 1984	13:45:46.6	18.70S	178.04W	6.0	422 I	462.6	A, C, L, NT, SD
Nov. 22, 1984	17:07:36.3	17.74S	178.05W	5.8	647 I	665.0	A, C, L
Mar. 15, 1985	00:16:01.6	20.61S	178.20W	5.6	526 I	568.0	A, C, J, L
Apr. 27, 1985	10:11:42.4	21.08S	176.81W	5.7	259 I	255.6	A, C, L, SD
Aug. 24, 1985	06:53:15.7	22.01S	177.79W	5.5	355 I	362.6	A, C, L
Aug. 28, 1985	20:50:49.0	21.00S	178.99W	6.1	629 I	646.6	A, C, J, L, SD
Sep. 15, 1985	11:25:05.3	19.22S	175.60W	5.6	258 I	245.2	A, C, L, SD
Dec. 25, 1985	22:15:09.1	21.64S	178.52W	5.5	447 I	490.9	A, C, J, L
Feb. 20, 1986	12:16:41.8	22.06S	179.56W	5.7	602 N	615.2	A
Apr. 04, 1986	10:13:40.7	18.04S	178.54W	5.8	540 N	557.0	A, C, J, L, SD
Jun. 16, 1986	10:48:25.7	22.04S	178.93W	6.3	547 N	571.0	A, C, J, L, NT, SD
Jun. 28, 1986	05:03:47.4	20.04S	176.06W	6.1	211 N	225.9	A, C, J, L, NT, SD
Jul. 02, 1986	04:16:57.7	22.04S	179.53W	5.6	598 N	604.8	A, C, J, L

Station records used: A=ANMO, C=COL, J=JAS, L=LON, NT=RSNT, SD=RSSD.
Event information used: I=International Seismological Center, N=National Earthquake Info. Center.
Event depths: 'd'=catalogue depth, 'h'=depth used for this study.

Figure 8) Data records from this study plotted against earthquake depth. Amplitudes are equalized between the different traces. (a) Data from earthquakes with depths between 500-700 km. Records are aligned at the first arrival. (b) Data from earthquakes between 200-500 km depth, with records aligned at the first arrival. The depth phase pP is the most prominent feature after the first arrival in each record.

(a)


P-phase
690
659
628
597
566
535
Event depth (km)
0
51
102
153
204
255
time (sec)

(b)

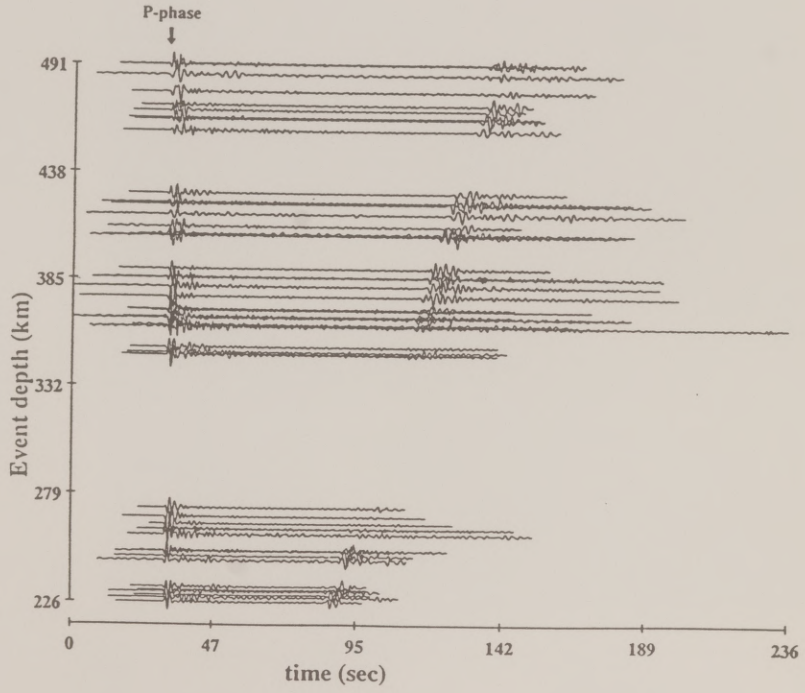
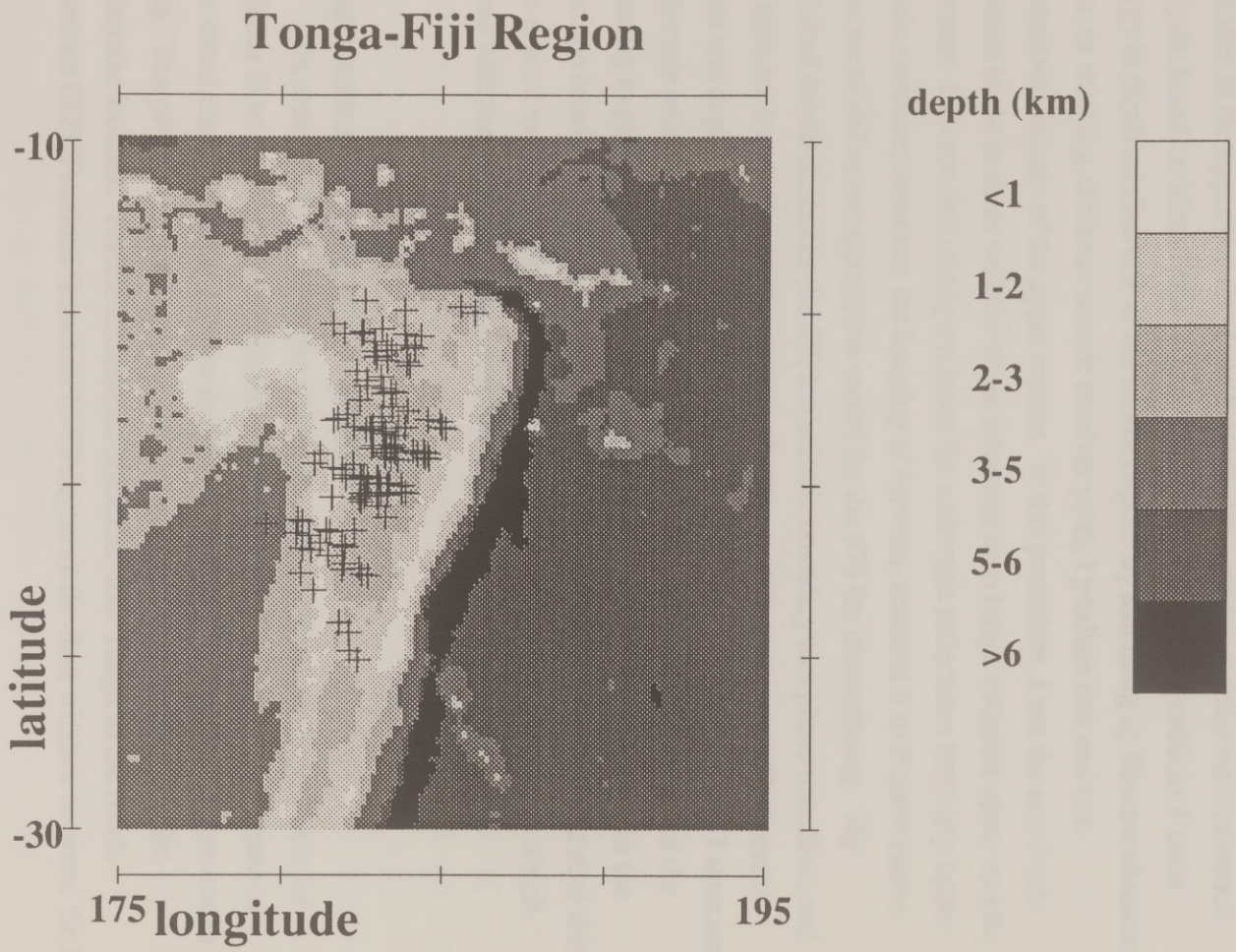

P-phase
491
438
385
332
279
226
Event depth (km)
0
47
95
142
189
236
time (sec)

Figure 9) The pP surface reflection points for the data records of Table 1, plotted with the bathymetry of the region. The concentration of earthquake sources, and the fairly common azimuth to the North American receivers makes the data coverage fairly dense, especially transverse to the trench axis.


Tonga-Fiji Region
depth (km)
<1
1-2
2-3
3-5
5-6
>6
-10
-30
latitude
175
longitude
195

above the earthquake (pdP, where d denotes the depth of the reflection), I can assume the scattering is from S-P conversions above the earthquake (sdP), and I can associate the coda energy with S-P forward-scattering in transmission below the earthquake depth (SdP). I start with an established model, PREM (Preliminary Reference Earth Model, Dziewonski and Anderson, 1981), for the background velocities in the upper mantle. I use the velocity model in a delay-time moveout analysis, based on ideas from seismic inverse theory, which positions P coda energy in depth relative to P at the source depth. By summing up the contributions from, or stacking, different source-receiver pairs, I produce one and two-dimensional images of the upper mantle. In this procedure, I use the amplitude variation in the coda as well the phase information to image coherent upper mantle structure. I do not find strong evidence for underside reflections from any upper mantle structure; however, the imaging of forward-scattered S to P conversions helps me to define an appropriate model for the 670 km discontinuity. My preferred model describes the 670 km seismic boundary as a phase transition, and I use iterative forward modeling to invert for the model parameters of horizontal extent and depth change of the boundary due to the presence of the slab. I associate the former model parameter with the slab width. I use the depth change of the boundary to give me a measurement of the pressure difference required for the transition relative to the change in temperature between the mantle and the cool slab. This gives a measurement of the Clapeyron slope for the phase transition which agrees with theoretical and experimental results.

THEORY

Upper mantle structure and the P wave coda

In order to study the upper mantle structure with seismology, I look for evidence from those phases which have sampled the upper mantle over some of their travel path. The time window between P and the pP phase is a unique one to study. The pP phase is a reflection off of the surface of the Earth above the earthquake location. For the teleseismic distances from Tonga to the North American GDSN stations (between 70-95 degrees), the time window between the P

wave arrival and the arrival of pP is uncluttered with other phases. For most of the deep events of Table 1, this time window is on the order of two minutes.

To study the upper mantle using the P wave coda, I choose large, deep earthquake events. Records for these events show little surface wave excitation and the P wave coda is less extended than for shallow events. Because these events occur in subduction zones, there is a question of whether any non-geometric arrivals, especially a path up the slab, could actually come in earlier than the pP phase. A quick calculation shows that an average velocity change of at least 20% between the slab and the surrounding mantle would be needed in order to bring this type of arrival in ahead of pP unless the slab is very steeply dipping. The difference in velocities in the 200-700 km depth range is only around 4% (Creager and Jordan, 1986). Since pP is a minimum time path, any scattering effects due to propagation along the path, such as interaction near the core-mantle boundary or as scattering in the crust and upper mantle (King et al.,1976), will come in later than the pP phase. Hence, any coherent arrivals prior to pP relate to seismic energy which has been scattered from the near source region.

Inverse theory and an imaging technique

Long wavelength features in a medium can be recovered through tomographic imaging of forward scattered energy. In order to get a more highly resolved structural image, shorter wavelengths and path lengths must be used and back-scattered energy should be considered if possible. Migration is a process by which energy in time is repositioned to energy in space by a mathematical treatment of the ray or wave equation. Whitcomb and Anderson (1970) performed a mapping of travel times to depth in a simple ray migration. They took travel lead times from P'P' and used a velocity model to calculate a depth of reflection for each lead time, plotting reflection points and smoothed histograms of the ensemble of reflections as functions of depth in order to map out reflecting boundaries. One mathematical treatment is Kirchhoff migration which is based on the symmetric form of Green's Theorem. This is basically a rigorous application of the Huygens' principle formulation that leads to the Kirchhoff integral, which is very similar to the 1960's diffraction summation approach (Schneider, 1978). While travel time inversions

use only phase information in the seismic record for determining a velocity model, migration uses both travel time information and amplitude information in order to image perturbations to the velocity model. Recent work has formalized this direct connection between migration and full waveform inverse theory (Tarantola, 1984; Mora, 1987; Docherty, 1991).

Following the notation and procedure of Cardimona (1991), I can write the linearized forward problem, i.e after applying the Born approximation, as

$$\delta \mathbf{u} = \mathbf{G}\, \delta \mathbf{m} , \qquad (1)$$

where $\mathbf{G}$ is the operator (Frechet derivative) that maps perturbations in model space ($\delta \mathbf{m}$) to perturbations in the data space ($\delta \mathbf{u}$). In the inverse problem a functional, or objective function, defined over the model space must be minimized (or maximized) in order to get the model which best fits the data. Assuming Gaussian statistics for the noise makes the problem the least squares one. Using a gradient method, Tarantola (1984) derives an iterative procedure to solve the non-linear inverse problem. The first iteration using the gradient method for a change in the model parameter vector is given by the adjoint $\mathbf{G}^*$ operator to the forward problem (Tarantola, 1987, Cardimona and Garmany, 1991), and the result is

$$\delta \mathbf{m}_1 = \mathbf{G}^* \, \delta \mathbf{u}_o . \qquad (2)$$

As shown by Cardimona and Garmany (1991), the application of the adjoint $\mathbf{G}^* = \mathbf{C}_m \mathbf{G}^t \mathbf{C}_u^{-1}$ to the data residuals includes the model space smoothing information implied by the covariance operators $\mathbf{C}_m$ and $\mathbf{C}_u$, which also describes any approximations included in the forward theory. After deriving $\mathbf{G}$ through the Born approximation of some wave equation perturbation problem, the operation of $\mathbf{G}^*$ on the data residuals involves using a background velocity model and the ray or wave equation to reposition energy in time to energy in image space (the resolving space for the model parameter), and a cross-correlation operation that gives an image density.

When looking for perturbations to a background velocity model, the image space is the one, two or three dimensional physical coordinate system, and we get a migration algorithm. Point by point in image space, an estimate of the source function is forward propagated from the source location to the image point, the data misfit is back propagated from the receiver to the image point and the resulting pair of time series are correlated to yield an image density. Numerically, the image density is proportional to the estimated changes required in the model parameter. The resolution of the model parameter is dependent upon the frequency content of the data. This is an important reason why in this study I chose to use short period data records, whereas most of the recent structure studies in global seismology have been restricted to long period data analyses.

This inversion analysis is dependent upon the linearization of the forward theory, given by the operator **G**. Hudson and Heritage (1981) have studied the general applicability of the Born approximation. Since they designed their analysis to encompass a wide variety of problems, it is not surprising their results show that the weak, single scattering assumption of the Born approximation may be inappropriate for many seismic scattering problems. With an appropriately defined model space, these inverse methods can be applied to global seismic data. Cardimona (1991) has shown that, for a one-dimensional linear model of the Earth (a reflectivity series), a gradient solution to the full waveform inversion problem reduces to the application of a matched filter (Treitel and Robinson, 1969). Revenaugh and Jordan (1987) used a matched filter analysis to invert ScS reverberation phases for mantle reflectivity structure. There is no limitation on the size of the region to be imaged, as long as the data coverage is sufficient for the analysis. McMechan and Fuis (1987) achieve outstanding results on some wide-angle reflection data of the U.S. Geological Survey from southern Alaska, showing that migration is capable of producing images which delineate the positions and geometry of tectonic features. Xie and Mitchell (1990) use a backprojection method similar to the solution of a full waveform imaging problem to map changes in coda Q from a constant background Q model. Questions of resolution and the smoothing in the model that would be required are dependent on what area of the Earth is being

imaged, and what the data coverage is in that area. The potential resolution is better than that of existing imaging techniques.

DATA PREPROCESSING

Short period global seismic data

I have concentrated in this chapter on using short period global seismic data to study the variability of local structure within the vicinity of the Tonga subduction zone. Short period seismic records have largely been ignored for structure studies apart from travel-time analyses because it is hard to get the alignment of phases down to within a half period for high frequency data. Because I am confident in my travel-time picks (to within .1 second) for the first arrival and the depth phase, I can get reliable depth estimates for my earthquake sources. Then my analysis is well constrained. I use travel-time differences relative to the first arrival to look for coherent structure relative to the source depth. However, to help reduce any phase alignment problems, I low-pass filtered the data at .8 Hz with a 3-pole Butterworth filter. Although this low-pass filtering puts the maximum information in the data on the low end of the full receiver instrument response, there is still substantial signal in the records at those frequencies.

The ray path and near receiver effects

The coda of the teleseismic P wave includes ray path and near receiver scattering effects as well as any near source phases. The near receiver effects are perhaps the worst; however, it is difficult to describe the receiver function using only vertical component data. Unfortunately, it has been only recently that short period data has become more widely available with three components of motion. The principal scattering mechanism which creates the near receiver effects in the coda may not be the broadband scattering seen in local earthquake codas (e.g. Aki, 1969), and a recent and thorough review article (Herraiz and Espinosa, 1987) makes no particular mention of P codas from deep earthquakes. Since I am interested in the near source structure in this chapter, any other scattered energy I consider to be noise. My analysis is defined so that, through stacking the data, I

may enhance any coherent near source phases at the expense of other scattered energy.

The source

From the original short period data (before low-pass filtering), I picked the arrival time of the P wave for each record, as well as a good arrival time for the pP depth phase, defined by the sea-floor reflection. Picking the pP arrival was harder, presumably due to the presence of crust and sea-surface phases. I obtained a best estimate depth for each event given by the pP-P times of records for that event. In general, the depths reported by the National Earthquake Information Center tend to be deeper than those of the ISC, and my inverted event depths (relative to PREM, and shown as 'h' compared to the catalogue depths 'd' in Table 1) were also deeper than the ISC reported depths. These differences are presumably due to different velocity models for the upper mantle.

The inverse problem for structure is complicated in part by the lack of knowledge about source processes for the deep earthquakes. The hypothesis of an isotropic source for deep events is in general not valid. Deep earthquake sources are predominantly deviatoric, and the description of these sources must be more complex (Frohlich, 1988). However, in the short period records for the events in Table 1, the first arrival P wave signals are not nodal and look simple It is clear that the source processes themselves must be fairly simple. Pennington and Isacks (1979) did a waveform analysis of short period signals from a set of deep earthquakes in the Tonga subduction zone, finding that they could use a similar simple source time displacement pulse to describe the waveforms from 60 of the 70 different events they considered. The earthquake sources are in general down-dip compressional (Giardini and Woodhouse, 1984). Wicks and Richards (1991) have correlated the intermittent observation of the S670P phase with changing focal mechanism, where a near nodal P direction is most favorable for the occurrence of the scattered phase. I know that I should get, in general, both shear and compressional energy in the azimuth to the North American GDSN stations from the Tonga subduction zone.

Seismic source studies depend heavily upon knowledge of the structure. On the other hand, structural studies can be done without detailed source information. In this study, I assume the source-time functions of each of the earthquakes are relatively unknown. Because I am using records from many different events, I need to equalize the data records relative to the different source strengths prior to any structure analysis. After removing any linear trend and subtracting the mean value in each record, I normalized the amplitudes between the records by scaling each data trace to unit variance, i.e., scaling by the total energy (the sum of the squares of each of the data values) in the record.

To normalize the phase between the different records, I include into a 'source' wavelet all the different effects on the seismic wavelet that I want to remove, including the source radiation pattern and polarity, any effects due to transmission through the Earth such as near receiver wavelet effects, and the instrument response itself. I applied a wavelet processing step to each data trace, using the autocorrelation of the first arrival in each record to get an estimate of the minimum phase wavelet which describes the 'source' signature (Yilmaz, 1987). I then cross-correlated each source wavelet estimate with each corresponding trace. This allowed me to make all the data records close to zero phase, with the polarity variation relative to a positive spike for the P wave. This procedure comes directly from inverse theory (Cardimona, 1991), and is associated with obtaining a one-dimensional Earth model of a reflectivity series in time (Oldenburg, 1981). The cross-correlation implies that our desired output is to be a maximum when the phase of the data is most like the filtering wavelet. The match filter operation is the best filter for achieving this in the presence of abundant noise (Treitel and Robinson, 1969).

IMAGING DATA ANALYSIS

Vertical delay-time equations

The simplest migration of a data set consists of mapping the time axis of a seismic trace to a depth coordinate, the mapping being derived from a given stratified velocity model. The velocity model gives the correct time moveout of the data, and the assumed starting model is one without any discontinuities in

structure. With a given velocity model, the pP phase gives us a reliable depth estimate for each earthquake. For imaging, I associate the P wave arrival time with the earthquake depth z=h in model space. I map the energy after P to depths z around the earthquake location using the background velocity model. I have taken the earth-flattening approximation (Aki and Richards, 1980) of PREM velocities as the background velocity model to get u(z) and w(z), the P wave and S wave slowness versus depth functions for the upper mantle, respectively. With these functions, the time after P of arrival of the reflection event from depth 'd' above the earthquake depth is given by (for ray parameter p)

$$pdP - P = 2 \int_{d}^{h} dz \left(u^2(z) - p^2 \right)^{1/2} . \qquad (3)$$

Here, the integrand is twice the vertical P wave slowness, u_p , at depth 'z'. The time after P at which an s to P conversion event from depth 'd' above the earthquake arrives is given by

$$sdP - P = \int_{d}^{h} dz \left(u_s + u_p \right) . \qquad (4)$$

In equation 4, the integrand is made up of the sum of the S wave vertical slowness and that of the P wave. For shear energy traveling down from the earthquake and converting to compressional energy, the time after P at which this scattered energy arrives from below the source is given by

$$SdP - P = \int_{h}^{d} dz \left(u_s - u_p \right) . \qquad (5)$$

The integrand in equation 5 is the difference between the S wave vertical slowness and that of the P wave. Equations 3-5 are approximations in that I have assumed

the ray parameter is the same for the upper mantle phases as for the direct P wave. Although this is not exactly true, the time error I introduce with this assumption for the ray parameter aperture I am dealing with and the teleseismic distance between the sources and receivers is at least an order of magnitude smaller than the phase picking error itself.

Equations 3-5 give the vertical delay-time moveout relationships for mapping energy in the P coda to upper mantle depths 'd' relative to zero depth at the sea floor and earthquake source depth 'h'. In performing this mapping from the time domain to depth space, the time series is moved out differently using each of equations 3-5. The result is three different normal moveout (NMO) analyses. After applying the moveout corrections given by either of equations 3-5, I then have each record mapped to depth above (equations 3 and 4) or below (equation 5) the earthquake sources. Since the preprocessing was meant to normalize the individual data records, I can stack the records after the delay-time migration into a one-dimensional depth series. I use the source location and azimuth to receiver information to bin and stack subsets of the individual depth records into two-dimensional images perpendicular to the trench axis. For the earthquake data applications, I suppressed the impulsive, high amplitude P arrival with a cosine taper on the depth-mapped data records before summing the series into an image. The cosine taper window extended to 120 km from the source position, which relates to a near-P taper covering about 10-20 seconds after the first arrival. This reduced even more any adverse effects due to uncertain source complexity.

Using traveltime differences in equations 3-5 eliminates the need for any station time corrections. I put in a static shift for each trace, aligning the data as in Figure 8 to the P wave arrival, and then use differences in traveltime from the P arrival throughout the analysis. Another observation about equations 3-5 is that they depend on the background velocity model, in this case PREM, for determining the stretching applied to the data records. It is important to note that the particular choice of background model does not affect the inverse problem drastically, as long as it describes the smooth variations in the upper mantle. Upper mantle velocity models in general agree to within 2%. The Jeffreys-Bullen model, for instance, is about 1.5% slower than PREM. I can use the background velocity model to image

boundaries; but then the positions of these boundaries will be dependent upon that velocity model. This assumes that I use the same background velocity model to do the imaging as I used for determining the earthquake source depths. I incur greater absolute errors in the parts of the image farthest from the source location, since the effects of my imperfect knowledge of the velocity structure may be biased and cumulative. However, because I know that P maps to the source depth, errors in the location of the deeper structure remain small even with gross errors in the velocity model. The method of migration I have outlined is quite stable and is appropriate for the frequency content and distribution of the data of Table 1.

Figure 10 gives an example of the steps involved in this data analysis. In the top of this figure is one of the data records from a deep earthquake event. The first arrival is quite clear, as is the pP phase about 2 minutes after the P phase. From the autocorrelation of the first arrival I define a minimum phase source wavelet estimate for the record. Cross-correlating this wavelet with the data record gives a record in time that is almost zero phase. With the example in Figure 10, I then use equation 3 above to map the record in time to depth in order to get a 1-D P wave reflectivity structure map above the earthquake depth. The P phase maps to the earthquake depth, and the pP phase maps to zero depth (the sea floor). I can use equation 4 to get an S to P reflectivity structure map, or equation 5 to get an S to P transmitivity structure map as well. The steps of cross-correlation with a source estimate and application of equation 3, 4, or 5 coincides with the action of the transpose operator in equation 2. The results from each record are summed together to build up a one or two dimensional image of the upper mantle.

Synthetic example

I calculated ray theory synthetic seismograms using PREM for upper mantle velocities, arbitrarily putting discontinuities at 100 km, 450 km, and 650 km. In the synthetic construction, I included only correct arrivals for pdP reflection events from this structure in the P wave coda, along with some arrivals near P which were uncorrelated with any structure. Using this artificial data set, I applied the three separate moveout analyses implied by equations 3-5. As shown in the depth sections of Figure 11 (a)-(c), only the pdP 'inverse' analysis successfully maps the

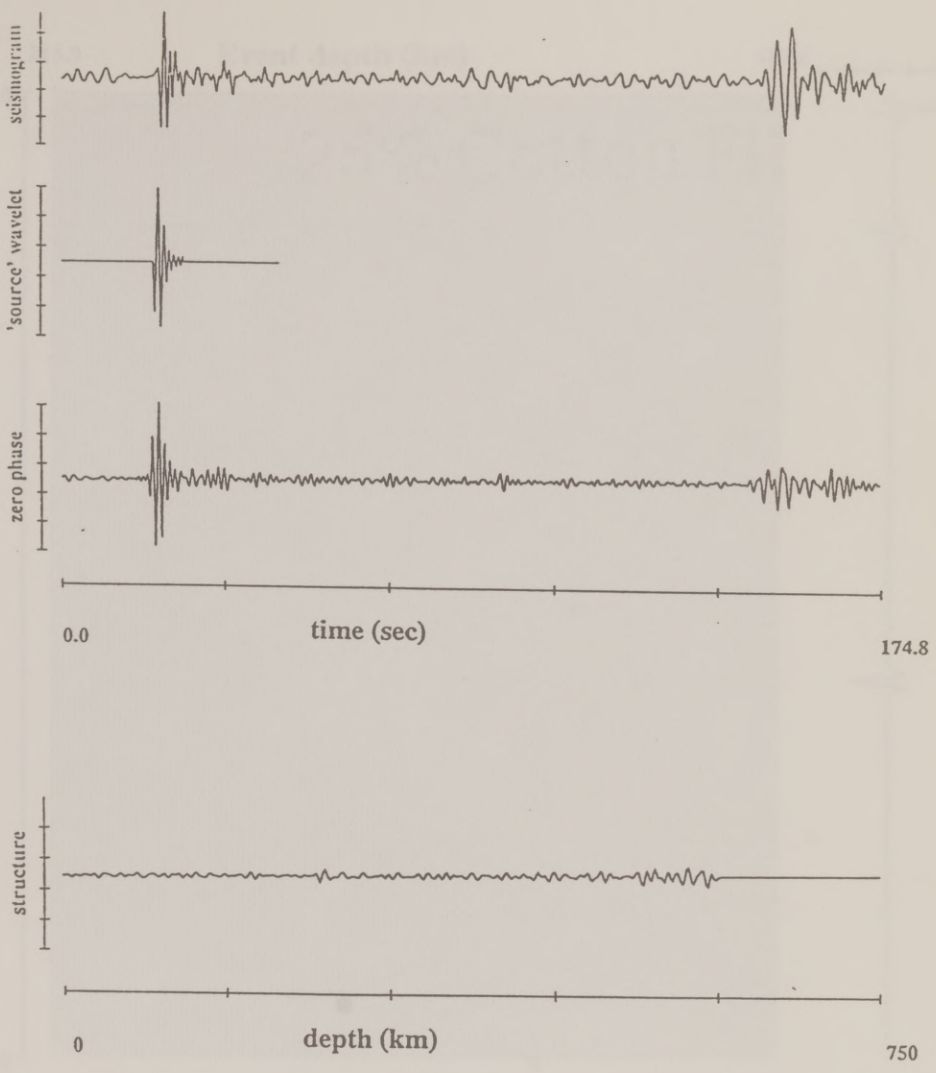


Figure 10) Example data analysis procedure. Top: a data record from one of the events of this study. From the auto correlation of the first arrival in the data record, I define a minimum phase source wavelet (mid-top) for the trace. Cross-correlating this wavelet with the record gives approximately a zero phase record in time (mid-bottom). Applying a moveout analysis with equation 3 maps the record in time to a record in depth (bottom). The P phase is mapped to the source location, and pP maps to the sea-floor.

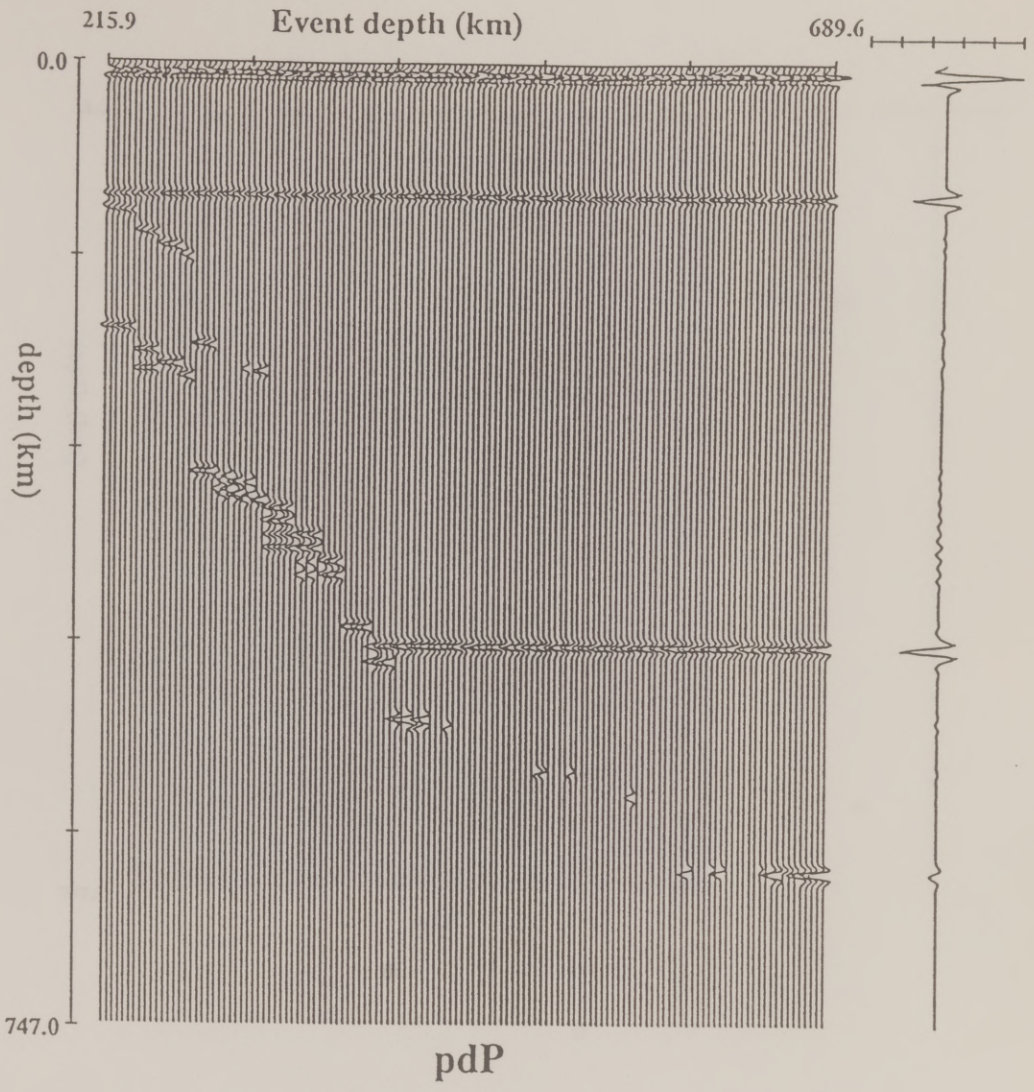


Figure 11 a) Depth section and single trace stack of the records after applying pdP data analysis (equation 3) to synthetics calculated to include only P reflections from artificial upper mantle structure. The correctly positioned scattered energy in each trace stacks out as it should when all records are summed together.

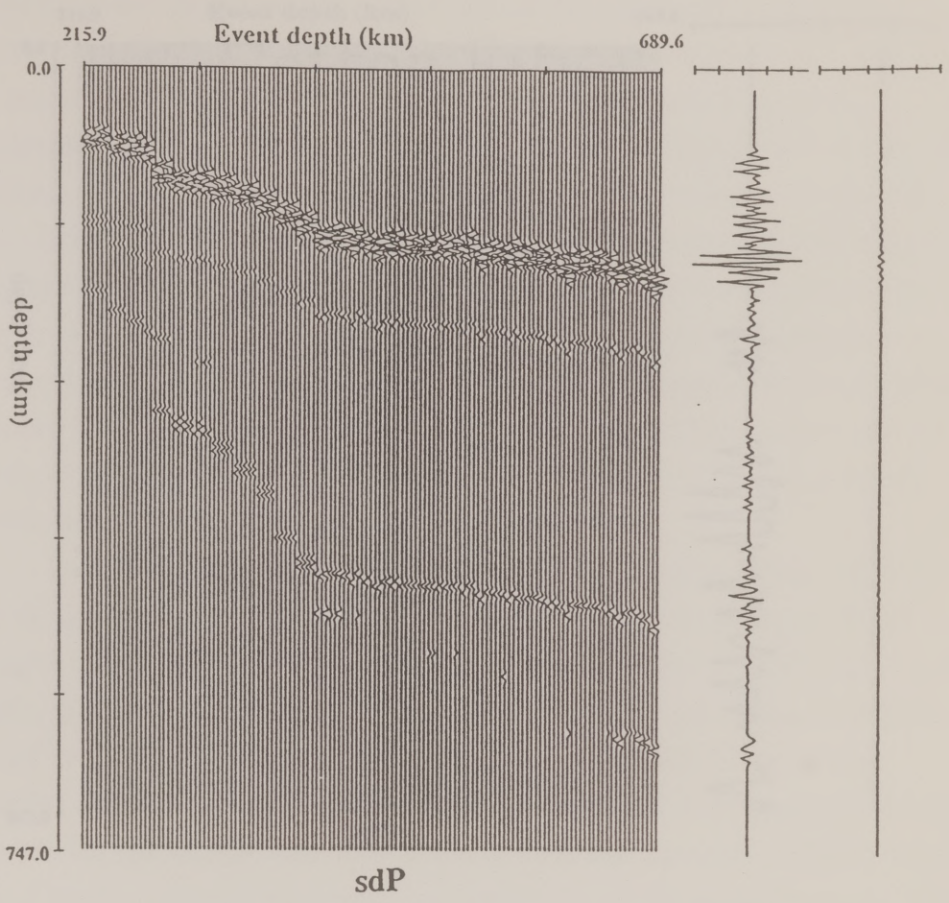


Figure 11 b) Depth section and single trace stack of the records after applying sdP data analysis (equation 4) to synthetics calculated to include only P reflections from artificial upper mantle structure. Single trace stack plotted to pdP stack scale at far right. The incorrectly positioned scattered energy in each trace does not stack out when all records are summed together. The pP surface reflection and the other reflected energy is considered noise in this analysis.

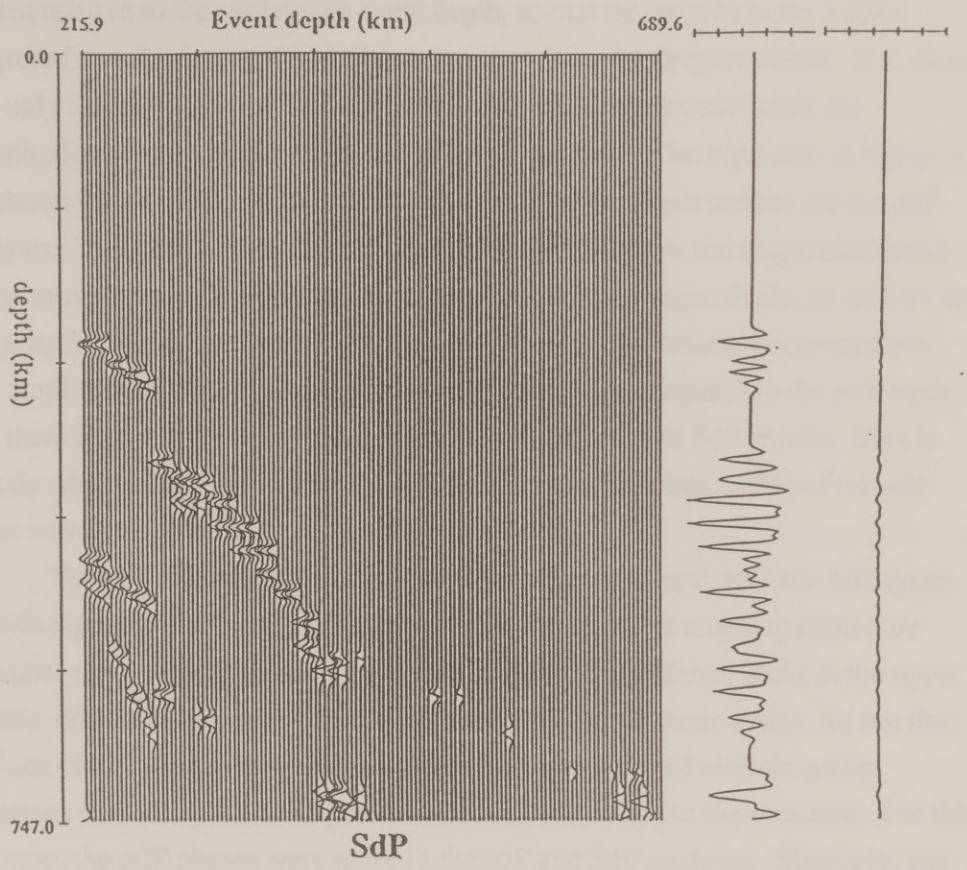


Figure 11 c) Depth section and single trace stack of the records after applying SdP data analysis (equation 5) to synthetics calculated to include only P reflections from artificial upper mantle structure along with some near P arrivals uncorrelated with any structure. Single trace stack plotted to pdP stack scale at far right. The incorrectly positioned pdP scattered energy in each trace does not stack out when all records are summed together.

coda energy to the correct positions in depth. In Figure 11, the records in depth are plotted relative to the earthquake event depth, so that the records in the section progress from the data of the shallowest to data from the deepest events. It is clear that only the pdP information is coherent, and hence 'stacks out' when the contributions from all the records are summed together. The right side of Figure 11 (a) shows the single trace stack of all the records in the depth section for the pdP analysis. The right side of Figures 11 (b) and (c) also show the single trace stack of the moved-out records for the sdP and SdP analyses, respectively, as well as the same stack in each case scaled to the amplitude of the pdP stack for comparison. The amplitude of the sdP and SdP stacks are near zero compared to the pdP stack, and there is obviously no model information in the sdP and SdP results. This is exactly what is expected, since the forward synthetic calculation did not include shear wave conversions.

The real data records may have information from pdP, sdP and SdP types of coda signals. The synthetic test shows that the moveout mapping procedure separates these three types of information, giving three different looks at the upper mantle. Since there was no sdP or SdP energy in the synthetic codas, we see that the lack of anything in the resulting stacks for these types of arrivals, using equations 4 and 5, gives us important null information about the structure. For this test case, the pdP phases were noise in the sdP and SdP analyses. Similarly, any other noise due to source complexity or scattering along the ray path to the receiver, such as near the core-mantle boundary, is noise in my upper mantle imaging and will not stack out when summing the data contributions into 1-D and 2-D images.

I can follow the individual ray paths through the upper mantle from the earthquake sources toward the receivers in North America. This allows me to bin the data so I can sum the individual record contributions into two-dimensional images of the upper mantle. The image plane is made up of 25 vertical bin locations. I project the ray path onto a two-dimensional image plane perpendicular to the trench axis, and stack the data into the image along this trajectory. I then apply a small amount of horizontal smoothing to damp the imaging. The smoothing is just a moving average filter across 3 bins at a time at each depth in the image.

Figure 12 shows the seismicity of events in Table 1 projected onto the image plane. The darker the plot, the more station records for that event.

RESULTS AND DISCUSSION

Imaging the free surface

Figure 13 shows the imaging results with a window around the pP depth phase and using the pdP analysis to get surface structure. The sea floor reflection is the strongest event, but the sea surface can be distinguished as well, at zero depth in the image. The imaging of the sea surface is not perfect, and that is due primarily to the variation in the bathymetry parallel to the trench. Since the water velocity is so low, the wavelength of the data in depth is small (about 1.5 km) which makes it hard to get coherent stacking in the image perpendicular to the trench. Nevertheless, Figure 13 gives a good indication of what this kind of imaging analysis can achieve. In the results to follow, I define zero depth in the images to be associated with the sea floor reflection.

Imaging the 670 km discontinuity with the S to P conversions

Using equations 3-5, Figure 14 shows the single trace stack of all the moved-out data for each of the delay-time analyses. The strongest imaging comes from the S to P conversion at the 670 km discontinuity. I see the strong conversion event in depth at approximately 675 km (below the sea floor, relative to the PREM velocity model). Although I have used the variation in amplitude of the seismic signals as information in the imaging, I have not tried to model the absolute amplitude. However, without knowledge of such things as the source directivity, I can still take a qualitative look at this conversion spike. I calculated the amplitude scattering coefficients (Zoeppritz, 1919) with appropriate angles of incidence for downgoing S and P waves incident upon the 670 km discontinuity. I used PREM to give the velocity and density jumps across this boundary, which were on the order of 5% for the P wave velocity, 8% for that of the S wave, and an 11% jump in density. Figure 15 shows these results. The S to P transmission coefficient is negative and is the same order of magnitude as all the other converted phases. For my ray parameter window, corresponding to distances of between 70 and 95

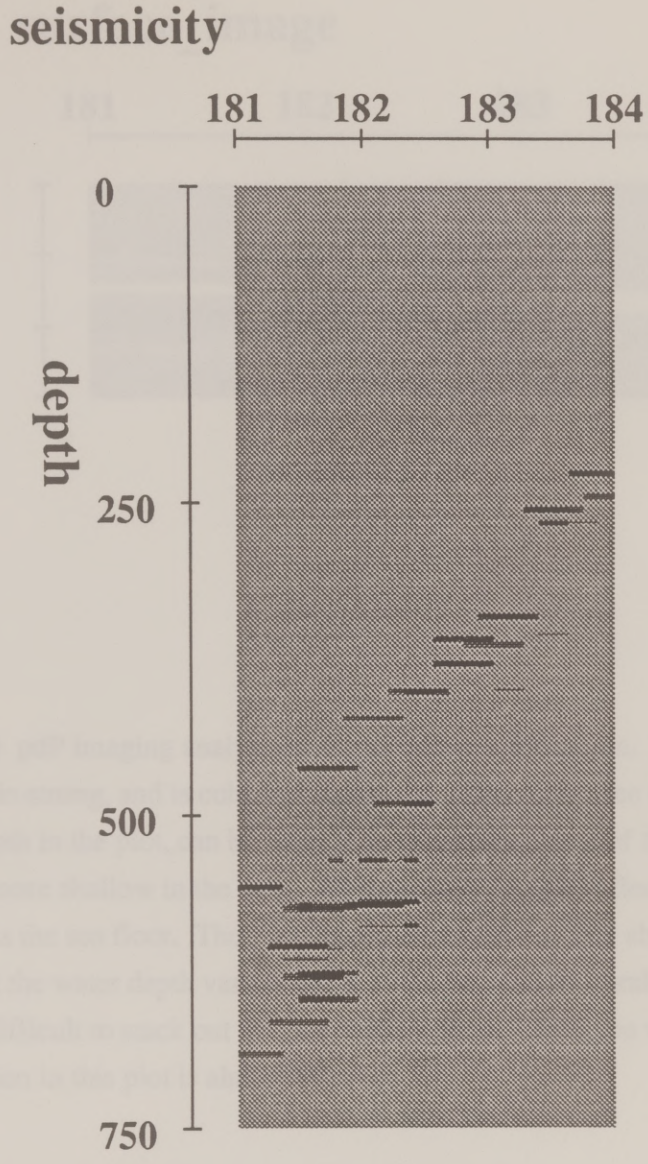


Figure 12) Seismicity pattern of events in Table 1 projected to an image plane perpendicular to the slab and the trench axis. Horizontal scale is East longitude. The darker marks imply more records for the event. Each source point is smoothed laterally by a small amount, which is the amount used to damp the imaging.

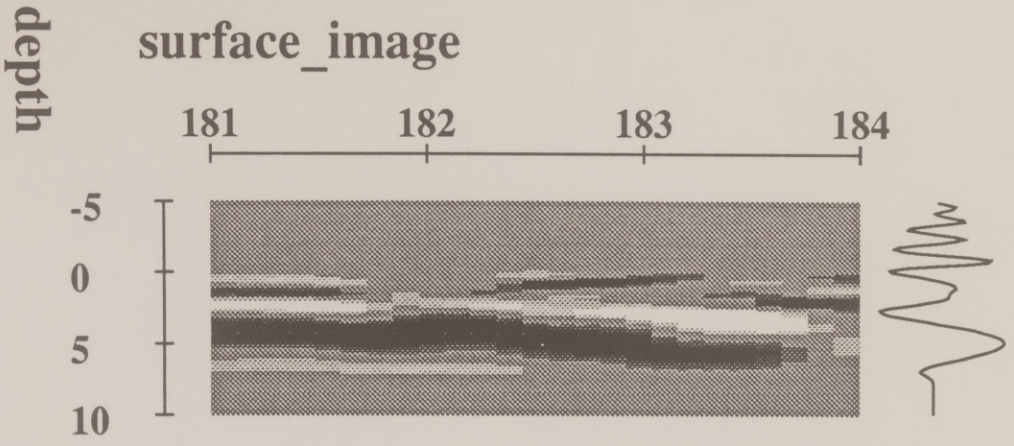


Figure 13) pdP imaging analysis of the pP phase information. The sea floor reflection is strong, and is coherent across the image. The free surface reflection, at zero depth in the plot, can be seen. The general structure of the water layer is evident: more shallow in the west. However, the surface reflection is not as coherent as the sea floor. The wavelength of the data is only about 1.5 km in the water, and the water depth variation due to the bathymetry parallel to the trench makes it difficult to stack out the free surface phase. Note, the vertical exaggeration in this plot is almost 10 times.

Figure 14) Single trace stack of all depth-mapped data records for each delay-time analysis given by equations (3) - (5). The only 1-D structure that seems strong in the results is at the 675 km depth from the SdP mapping.

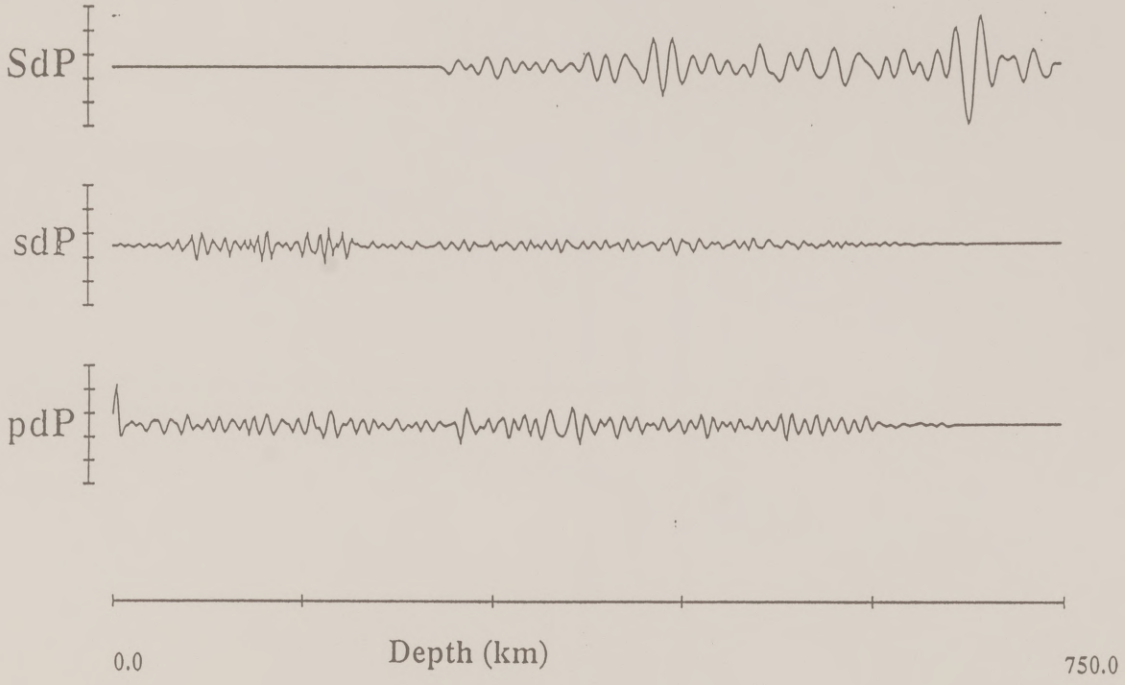
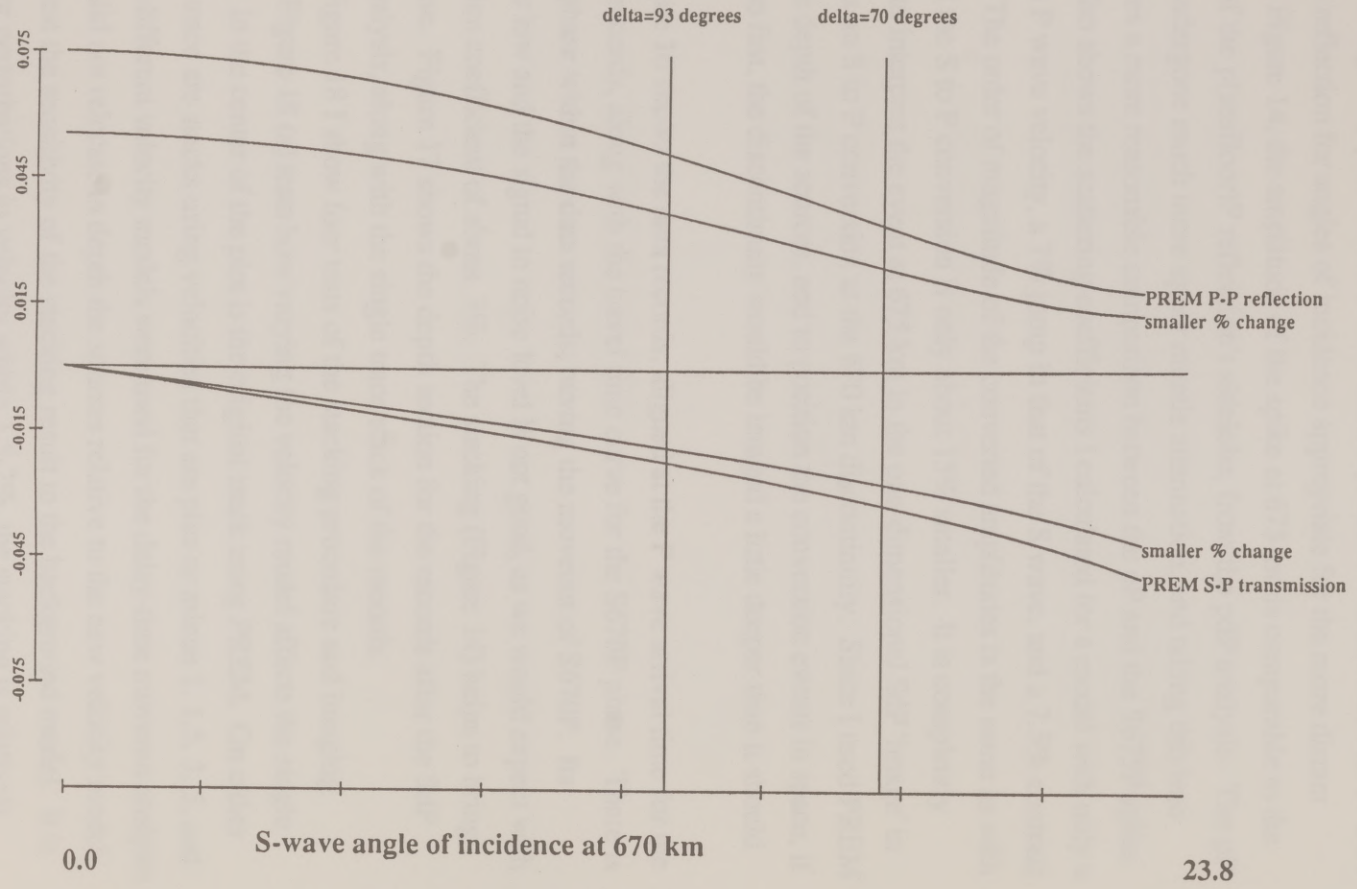

SdP
sdP
pdP
0.0
Depth (km)
750.0

Figure 15) Scattering coefficients for P to P reflection and S to P transmission at the 670 km discontinuity for PREM and a model with a smaller change in properties across the boundary. For the ray parameter window corresponding to distances of 70-93 degrees, the S to P conversion should be about 2-3%.

PREM scattering coefficients



degrees, the amplitude of S670P is about equal to the P to P topside reflection coefficient at larger ray parameters (closer stations), and it is about half that of the P to P topside reflection for angles of incidence appropriate for the more distant stations. In Figure 14, the amplitude of the spike at 675 km is comparable to the amplitude of the p(seafloor)P reflection's sidelobe, from the pdP analysis. The pP phase has undergone much more upper mantle attenuation, and taking this into account gives a more reasonable comparison between the pP and the S675P spike. Figure 15 also shows the scattering coefficients I calculated for a model with only a 4% jump in P wave velocity, a 7% jump in that of the S wave, and a 7.5% contrast in density. The order of magnitude of the converted amplitudes is the same as with PREM, and the S to P conversion is only about 15% smaller. It is completely reasonable to interpret the event at 675 km in the one-dimensional SdP 'image' in Figure 14 as an S to P conversion at the 670 km discontinuity. Since I used PREM to locate the depth of the sources, and to position the conversion events in space, if PREM is too fast, the discontinuity would be imaged a little deeper than it should be.

Figure 16 shows the data records, aligned at the P wave arrival time for the zero-phase records, along with the travel time curve for the S670P phase. There is a coherent phase within the data records, having the moveout of S670P. Its amplitude is low and the signal to noise level is not good, as we would expect with a transmission coefficient of about 3%. The stacking (Figure 14) helps to bring out this phase. Figure 17 shows the depth section for the records after the SdP moveout analysis, along with the single trace stack of the records.

In Figure 18 I show four tests of the stacking procedure and imaging technique. Figure 18 (a) tests how varying the velocity model affects the single trace stack. In the center of the plot is the original stack using PREM. On either side of this trace are stacks using velocities that are plus or minus 1, 1.5, 3, 5, and 7%. These different velocity models were used for the delay-time moveout analysis only, and I did not relocate in depth the sources relative to the new velocity model in order to test the sensitivity of the stacking result to the background model. It is clear that, for perturbations in velocity within 1.5-2%, the stacking is relatively insensitive to changes in the velocities. Of course, given a different background

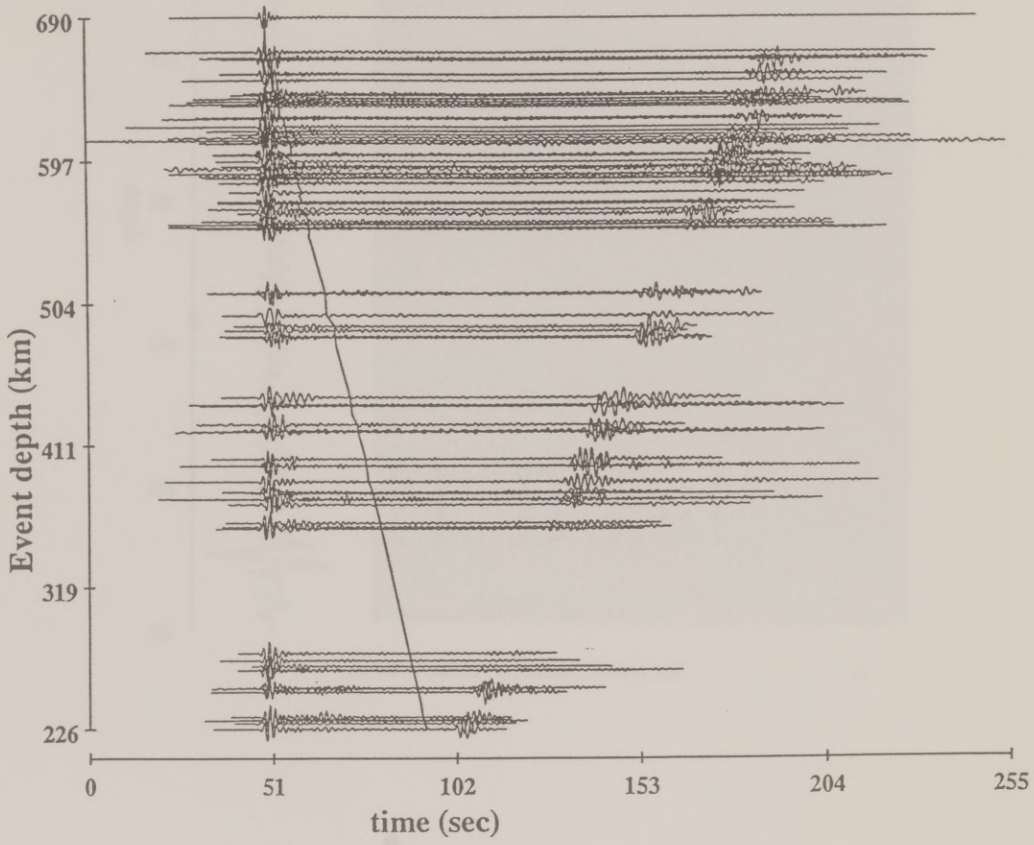


Figure 16) Zero phase data records aligned at the P-wave arrival time for each record, and plotted against event depth. The S670P phase travel-time curve is plotted as well. There seems to be noticeable coherent energy following the S670P moveout, although its amplitude is comparable to the noise level.

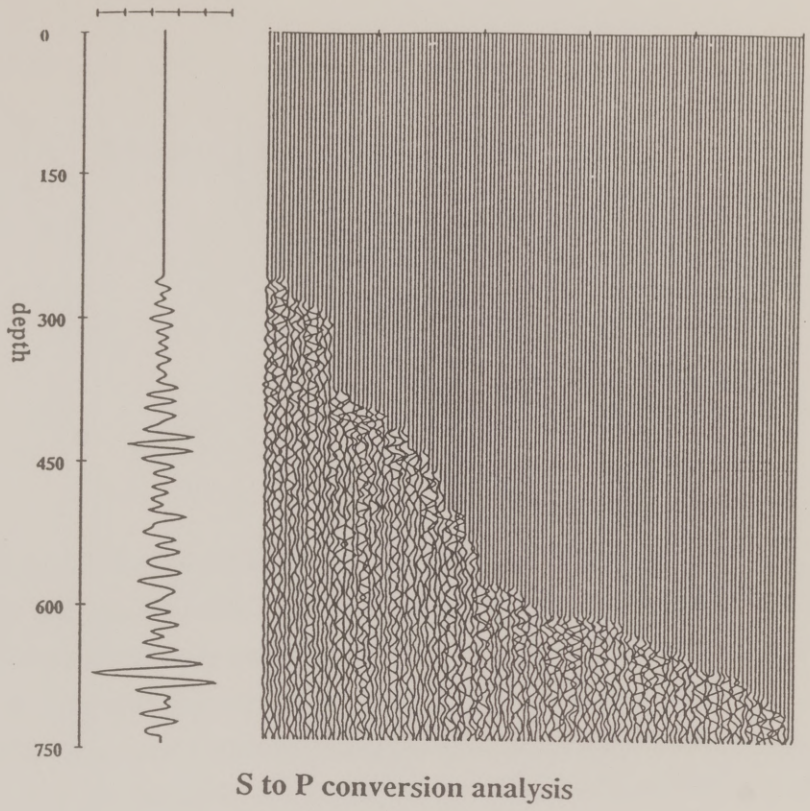


Figure 17) Depth section and single trace stack after applying the SdP moveout analysis to the data. The traces are equally spaced and in order of increasing source depth from left to right.

Figure 18 a) Test of sensitivity of stacking procedure to background velocity model. Without relocating the sources in depth, I plot the single trace stacks after applying the moveout analysis with velocity models perturbed from that of PREM by +/- 1, 1.5, 3, 5, and 7%. The original stacked trace is plotted in the middle at 0% perturbation. Most upper mantle velocity models agree within 2%, and the stacking procedure is relatively unaffected for this small perturbation in velocity. If I start with a different velocity model, to be consistent I would begin the procedure by relocating the source depths relative to the new model.

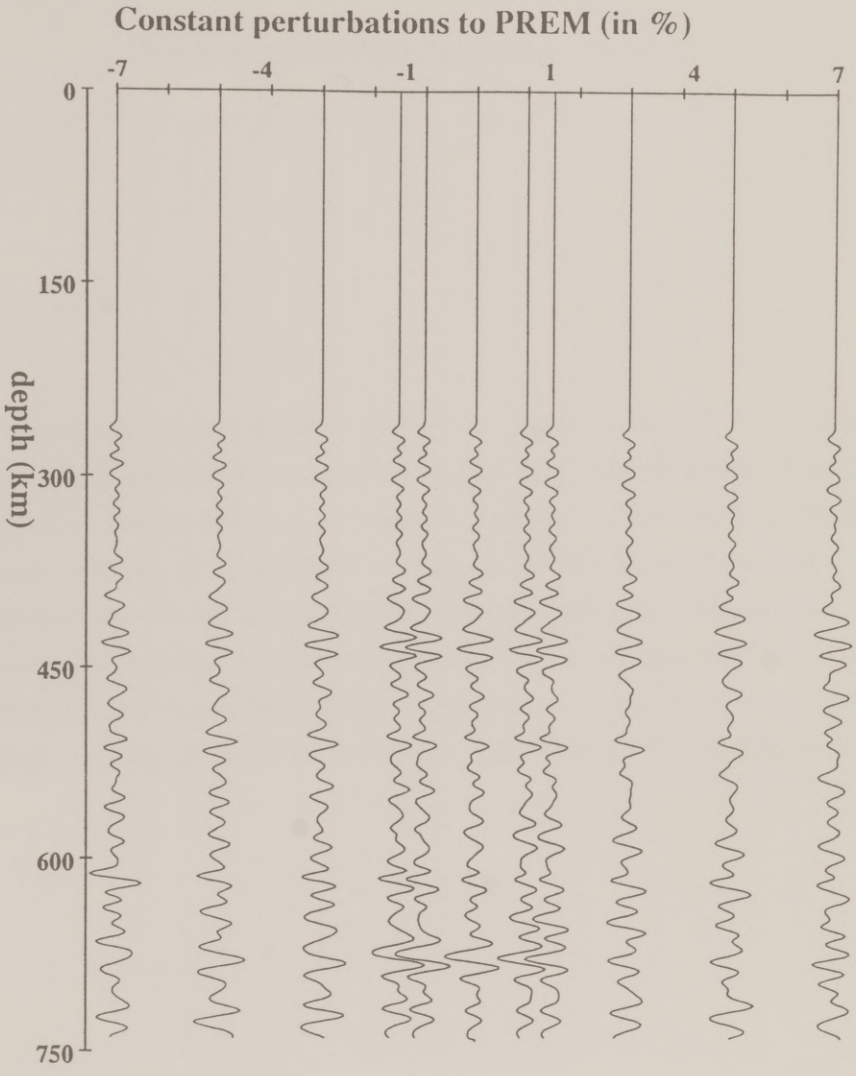

Constant perturbations to PREM (in %)
-7
-4
-1
1
4
7
0
150
300
450
600
750
depth (km)

Figure 18 b) Test of sensitivity of stacking procedure to mislocations in depth of the earthquakes. In this figure I add a random amount of time shift to each trace before stack which corresponds to shifting the records in depth, or having an incorrect depth for the earthquake source. This is a similar problem as in Figure 18 (a). For zero depth perturbation in this figure, we get the original single trace stack. As I increase the amount of random time shifts added to each trace, the stack looses coherency. The stacking/imaging method assumes the depths are very well constrained by the pP-P time, otherwise the procedure is meaningless.

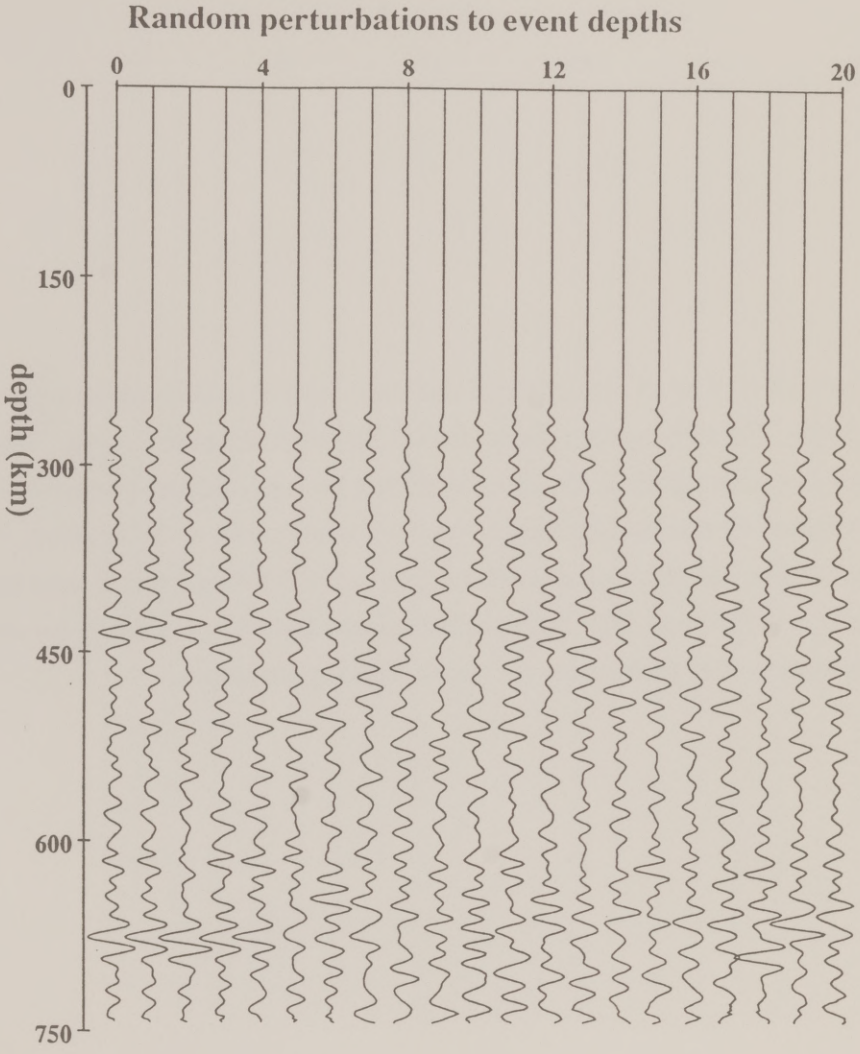

Random perturbations to event depths
0
4
8
12
16
20
0
150
300
450
600
750
depth (km)

Figure 18 c) Test of sensitivity of stacking result to random phase rotations in the records. The zero random phase stack is the original stacked result. Whether the stack contains signal or not, this plot shows that the stacking procedure itself is unaffected by random rotations in separate traces up to +/- 90 degrees. This is intuitively clear, but also shows how insensitive the imaging is to reasonable changes in the input data; in particular, related to the inaccuracy of the zero-phase assumption after performing the match filtering procedure.

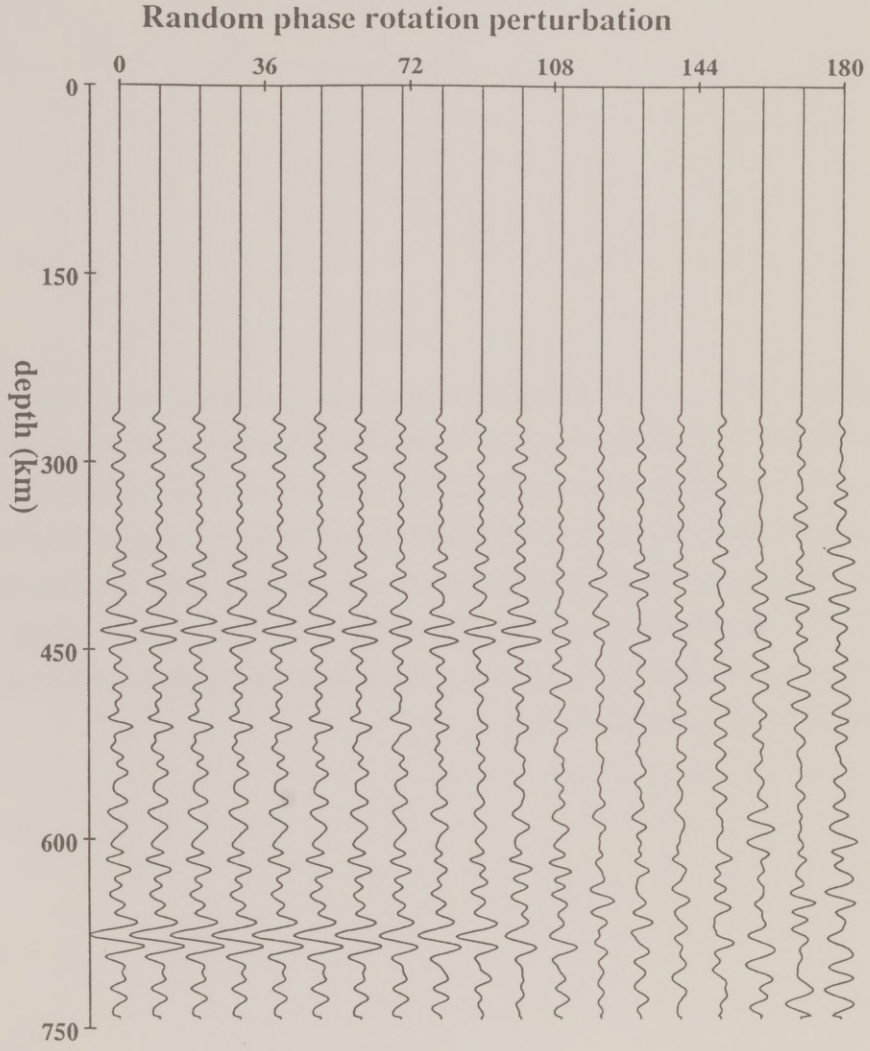
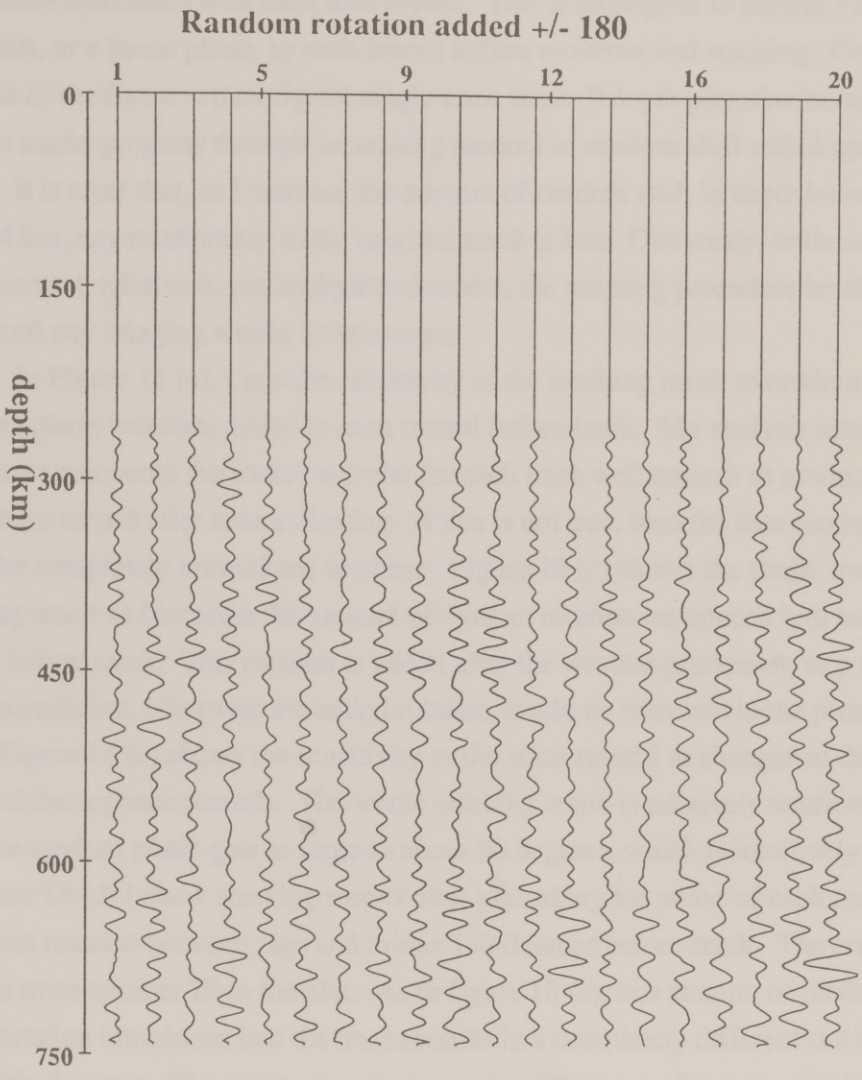

Random phase rotation perturbation
0
36
72
108
144
180
0
150
300
450
600
750
depth (km)

Figure 18 d) Comparison of the original stacking result with numerous realizations of randomized stacking results. The original is trace number 20. Each of the other traces is the sum of all the input data after adding +/-180 degrees of random phase rotation to each record before stack. As seen in Figure 18 (c), this amount of random phase rotation turns the data into noise. It's clear that the S675P phase in the original stack is on the order of the noise level since its transmission coefficient is only 2-3%, yet it does still stand out among the randomized traces in this plot.


Random rotation added +/- 180
1
5
9
12
16
20
0
150
300
450
600
750
depth (km)

velocity model, to be consistent the analysis should actually start with getting a source depth relative to that model. I show a similar test result in Figure 18 (b) where I use PREM as the background model, but I randomly perturb the depths of the sources associated with each data record. This is analogous to adding a random time shift, or a linear phase, to each record before moveout and stacking. On the left side of the figure is the original single trace stack (0 depth perturbation added), and the stacks progress through increasing amount of random shift added up to +/- 20km. It is clear that, as I increase the amount of random shift in depth beyond about 4 km, any relationship to the original stack is lost. Obviously, without good depth control, relative to the background model, the stacking procedure breaks down and any imaging would be nonsense.

In Figure 18 (c), I test the sensitivity of the stacking result to random constant phase rotations added to each record before stack. My analysis assumes that I have recovered the source wavelet for each trace well enough to produce a zero phase record after match filtering. If this is not true, then the data records won't be completely normalized in phase. Figure 18 (c) shows the single trace stacking result as I increase the amount of random rotation introduced into each record before stack. This rotation is added after the wavelet processing step and cross-correlation, otherwise the added rotation would be removed in the procedure. Thus, Figure 18 (c) shows the sensitivity to the stack related to changes in the phase of the separate records. The whole stacking result is relatively unaffected until the random phase gets as large as about 90 degrees, which is intuitively clear. In Figure 18 (d) I show stacking results after perturbing the phase of each record by a random rotation between plus and minus 180 degrees before stack. The original stack is trace number 20 in the plot. As in figure 18 (c), this amount of random phase rotation introduced into the traces results in a completely different stack from the original output. The comparison between the different realizations of this random stacking in Figure 18 (d) shows that the amplitude of S675P is close to the noise level of the data records, which we already knew from the amplitude of the expected transmission coefficient calculated earlier. Being careful to recognize the fact that the S to P conversion signal is very subtle as seen in Figure 16, I note that

the stacking does help to bring this coherent phase out somewhat above the noise level.

We can learn more about the boundary at 675 km depth by looking at the spatial variation of the S to P transmitivity in a two-dimensional view. Figure 19 shows the result with an image plane perpendicular to the strike of the descending slab. The single trace stack is shown for comparison. Another feature at around 715 km appears strong in the 2-D image. Figure 20 (a) shows the imaging results using a zero phase modeling wavelet with ray theory synthetics, calculated to model the amplitude and phase variation of the data records in order to reproduce the strong features of the actual stacked data image of Figure 19 in the vicinity of the 675 km discontinuity. In this modeling I assumed the scattering was from two distinct interfaces, and I allowed each interface to have a small amount of vertical variability in order to model the strong features of the data records. The data image for the same depth window is shown in Figure 20 (b) for comparison.

With the given data coverage, I tried to model in Figure 20 the general features of the data image at around 700 km depth by assuming the existence of a second discontinuity deeper than the 670 which might account for the deeper scattered energy. Richards and Wicks (1990) noted a 'second' S to P conversion in their data set which they thought was possibly related to conversion at a deeper interface. From petrological data, it is not clear what type of seismic discontinuity we would introduce at a depth of around 715 km.

An alternate interpretation of the imaging results in Figure 19, also mentioned by Richards and Wicks (1990) to offer an explanation for there data, is that the interface at 715 km is the same as that at 675 km, namely the spinel to perovskite phase transition depressed within the cold slab interior. This phase transition is endothermic, implying a negative Clapeyron slope which describes the differential change in pressure with respect to a change in temperature. Since the slab interior is cooler than the surrounding mantle by a substantial amount, the negative Clapeyron slope requires that the phase boundary should proceed at a greater depth within the slab than the surrounding mantle.

The band-limited data are sensitive to more than just the geometrical conversion points on the interface describing the 670 km discontinuity. Diffraction

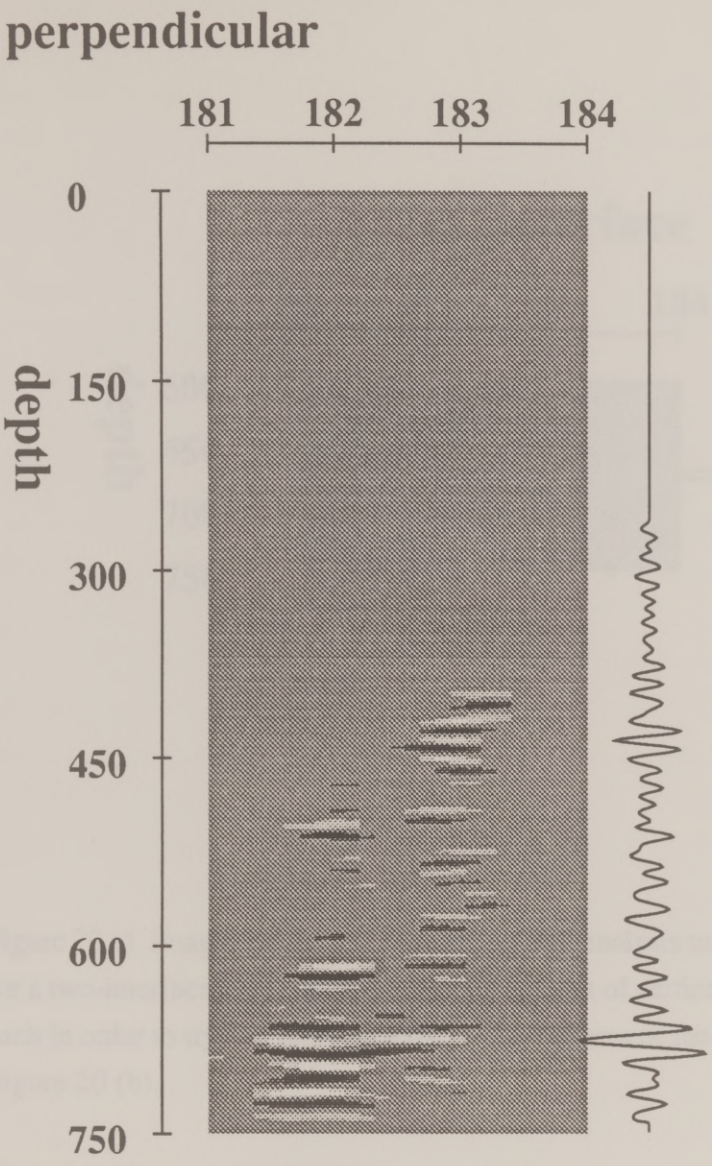


Figure 19) SdP moved-out data stacked into image plane perpendicular to the trench axis. The depth is in km, and the horizontal coordinate is East longitude. The single trace stack is plotted as well for comparison. Another structural feature besides that at 675 km comes out in this image at around 715 km depth.

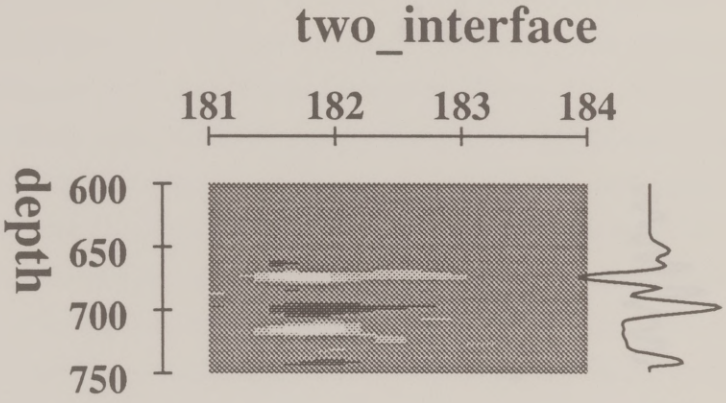


Figure 20 a) Imaging results after applying SdP analysis to synthetics calculated for a two-interface model with a moderate amount of vertical variability allowed on each in order to try and reproduce the strong features of the data imaging result of Figure 20 (b).

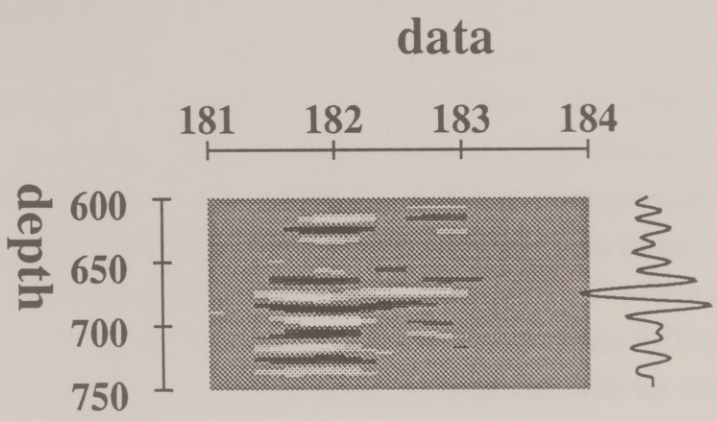


Figure 20 b) 600-750 km depth window of data imaging result shown in Figure 19. The synthetics imaged in Figure 20 (a) were calculated to try and represent the strong features in the data records in the region around 700 km.

theory formalizes this idea (e.g., Elmore and Heald, 1969). The seismic waves illuminate the interface, and excite converted energy via numerous diffracting points on the entire interface. These Huygens diffraction wavelets constructively interfere to create the forward and back scattered phases in the data. The wavelets constructively interfere around stationary phase points, i.e. points where the variation in phase is slow enough (stationary) that the peaks and troughs of the scattered components can add together constructively. For a flat interface, the stationary phase point would be the ray theory reflection or transmission point. The area of the interface actually contributing to the reflection or transmission is dependent upon the radius of the Fresnel zone, which gets larger as a (square-root) function of wavelength and distance from the interface. For deep events around 600 km, the Fresnel zone radius for incident S waves at the 670 km depth is around 30 km. Larger Fresnel zone radii would be associated with the shallower earthquakes.

When the illuminated interface is perturbed, the Fresnel zone may cover an area of lateral variation, and there may be more than one, or there may be no, stationary phase point around which the individual Huygens wavelets constructively interfere. It is physically probable that in Figure 19 I have imaged diffractions from the 670 km discontinuity both in the mantle and within the slab itself. Taking a simple model of a slab at around 670 km depth, parallel to the trench axis and positioned relative to the seismicity of the earthquakes of this study, I use Fresnel-Kirchhoff diffraction theory (Elmore and Heald, 1969) and a zero phase modeling wavelet to obtain synthetic seismograms for a single interface at 675 km depth that is perturbed, i.e. deeper, within the slab.

Inverting for the 670 km interface parameters

I can define objective functions to describe the correlation between the data and the results from synthetic calculations for a given model. I invert the data by iterative forward modeling to search the model space for a maximum (or minimum) in the objective function. For the model of a single perturbed interface, I introduce model parameters of slab width, or horizontal extent of interface perturbation, and change in depth of the interface within the slab. I calculated Kirchhoff synthetics

for 121 models, varying the slab width from 0.0 km (no slab expression, i.e. no depression of the interface) to 100 km, and the depression of the 675 km seismic boundary within the slab between 30 and 60 km. Figure 21 shows the plots of results for three different objective functions. In Figure 21 (a), I use a semblance, or coherency statistic. This objective function is always positive, and it tests the energy in the stack of the data and the synthetics, relative to the sum of energy in each separately. In Figure 21 (b) I plot a correlation-coefficient statistic, which shows the energy that is in phase between the data and the synthetic. Figure 21 (c) shows the results using a normalized squared-error (total error energy) objective function, defined so that zero error plots strongest. I seek the global maximum in each plot. The wavelength resolution in depth is evident in each plot from the Δz variation, but there is a well resolved peak at 41 km. The variation in slab width is more broad, approaching a constant value at a width of 0.0 km (no perturbation in the interface). There is still a clear maximum at 50 km. The same statistics as in Figure 21 (a) and (b) applied to the result using synthetics after modeling two layers in Figure 20 (a) are less than the global maxima of Figure 21.

Using the diffraction scattering from a single perturbed interface as my preferred model for the imaging results, and with the parameters obtained from the inversion through iterative forward modeling, I describe the 670 discontinuity in terms of a 50 km wide slab with a 41 km depression of the boundary within the slab. Imaging synthetics created for this model (Figure 22), I show that the diffraction model does indeed describe the qualitative features of amplitude and phase variation seen in Figure 20 (b). The single trace stack of these synthetics shows a strong peak at 675 km, and only a much smaller peak at 715 km, even though the 2-D imaging shows a bright spot as with the data result.

Figure 23 (a) shows an image after creating Kirchhoff diffraction synthetics for a 675 boundary that was depressed across a 60 km width, only about 10 km wider than for Figure 22. The strength of the imaging of the interface at 675 km is substantially reduced, and the 715 km depression becomes much too strong in the image compared with the real data results. In Figure 23 (b) I show the imaging results using Kirchhoff synthetics created from a model with a 40 km wide slab. In this image, the 675 km interface is too strong relative to the one at 715 km,

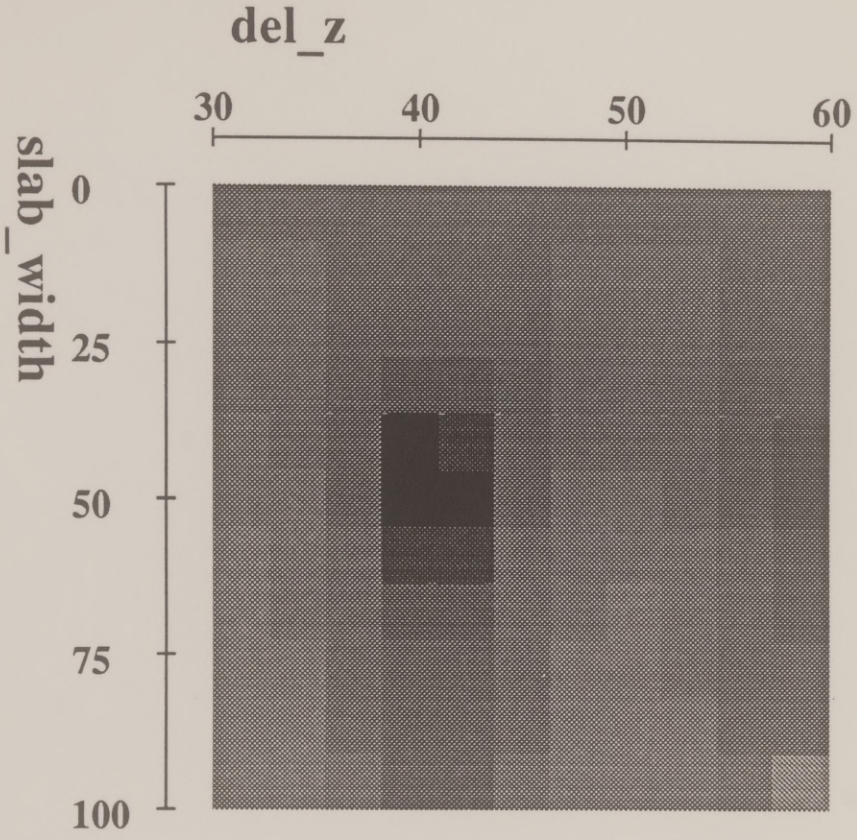


Figure 21 a) Objective function for the model space given by the semblance of the data with the synthetics, as a function of assumed slab width and the depth of the depressed interface relative to the 675 km deep discontinuity. This objective function is the ratio of the energy in the stack of the data with the synthetics to the energy in each separately. The global maximum is at the point in model space defined by a slab width of 50 km and a change in interface depth of 41 km.

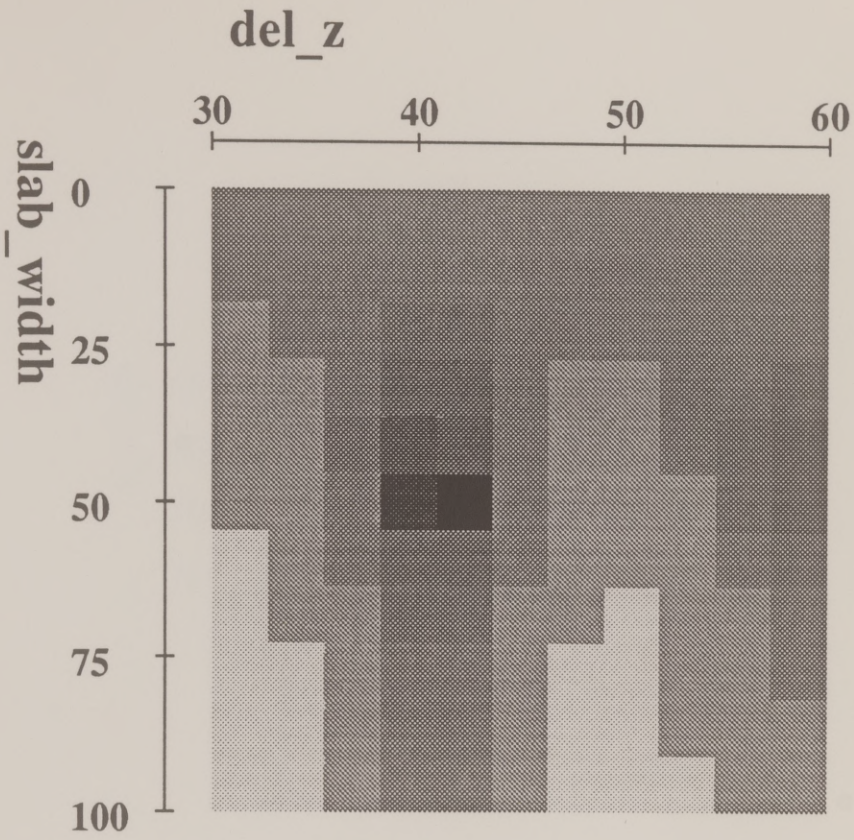


Figure 21 b) Objective function for the model space given by the correlation of the data with the synthetics, as a function of assumed slab width and the depth of the depressed interface relative to the 675 km deep discontinuity. This objective function is the ratio of the energy that is in phase between the data with the synthetics to the energy in the data alone. The global maximum is at the point in model space defined by a slab width of 50 km and a change in interface depth of 42 km.

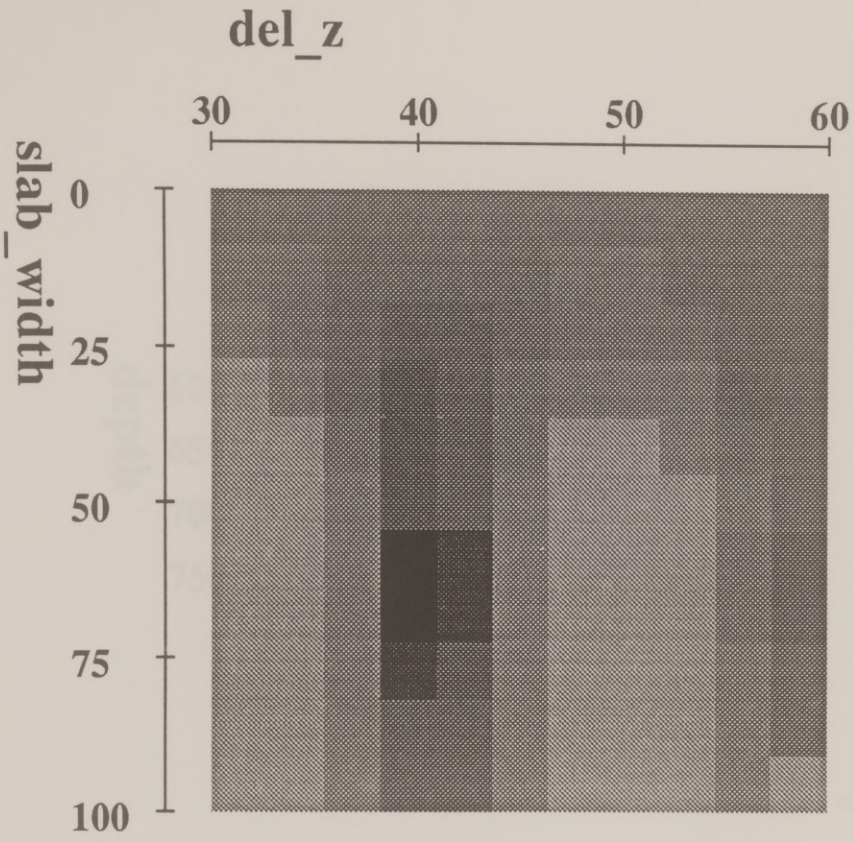


Figure 21 c) Objective function for the model space given by the squared error of the data minus the synthetics, as a function of assumed slab width and the depth of the depressed interface relative to the 675 km deep discontinuity. This objective function is related to the ratio of the error energy to the energy in the data alone, defined so that zero error would plot as a maximum. The global maximum is at the point in model space defined by a change in interface depth of about 40 km, and a slab width of about 60 km. This number is shifted a bit from the results of Figures 21 (a) and (b), but this statistic is much more sensitive to the amplitudes in the data and synthetics which are not modelled as well as is the phase variation.

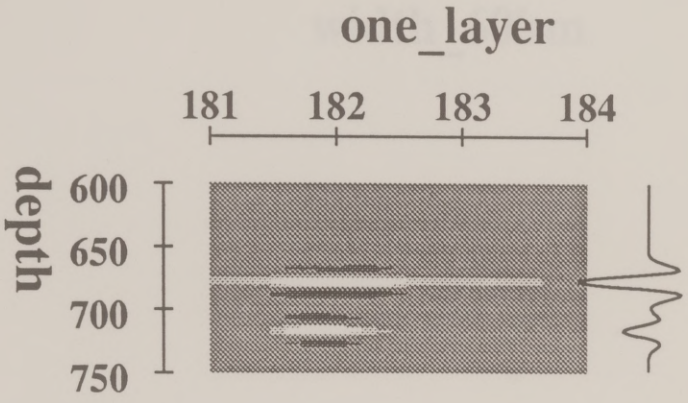


Figure 22) Imaging using synthetics calculated for a model of a single perturbed interface associated with the 670 km discontinuity. Slab width is taken to be 50 km, and the depression of the seismic boundary within the slab is taken to be 41 km. This model does a better job of reproducing the qualitative and quantitative features of the data results in Figure 20 (b) than did the two-layer model results of Figure 20 (a).

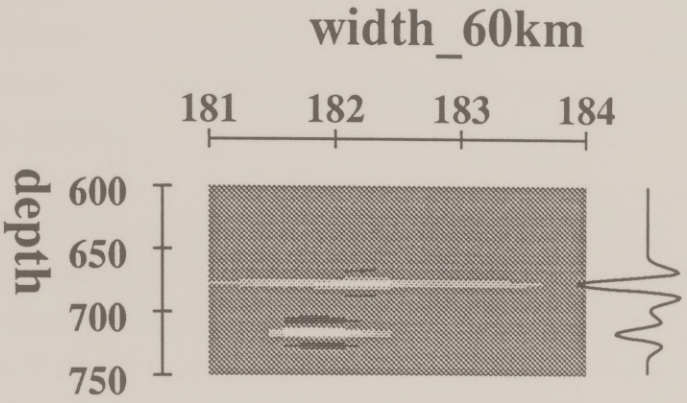


Figure 23 a) Imaging using synthetics calculated for a model of a single perturbed interface associated with the 670 km discontinuity. The depression of the seismic boundary within the slab is taken to be 41 km as in Figure 22; however, the slab width is taken to be 60 km. The structural imaging at 715 km depth is much too strong relative to that at the 675 km depth when comparing this result to the data result of Figure 20 (b).

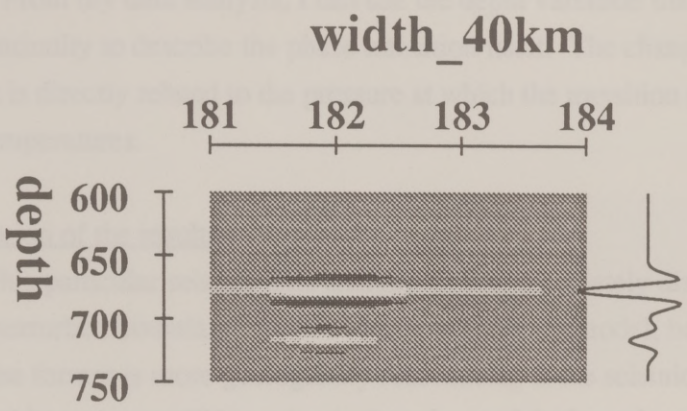


Figure 23 b) Imaging using synthetics calculated for a model of a single perturbed interface associated with the 670 km discontinuity. The depression of the seismic boundary within the slab is taken to be 41 km as in Figure 22; however, the slab width is taken to be only 40 km. The structural imaging at 715 km depth is much too weak relative to that at the 675 km depth when comparing this result to the data result of Figure 20 (b).

compared to the data results. Figure 24 shows the trace of the best model where it intersects the image plane, plotted with the data image of Figure 20 (b). The model interface coincides with the bright features of the image. From the iterative forward modeling results I can put limits on the width of the cool slab interior at the depth of 675 km. From my data analysis, I can use the depth variation that I find in the 675 km discontinuity to describe the phase transition itself. The change in depth of the transition is directly related to the pressure at which the transition takes place at mantle temperatures.

Interpretation of the results

This particular seismic data analysis cannot completely distinguish between a single perturbed boundary model and a two boundary model; however, I do assume the former is more geologically reasonable. If the seismic data are sensing a chemical boundary at 675 km depth, then the model of two-layer mantle convection implies that the slab would be restricted from penetrating into the lower mantle (Figure 25). The descending slab would presumably force a broad lateral perturbation on the boundary, and the depression would be large in depth as well, perhaps on the order of 50-200 km. The inversion results of the last section relate to this type of boundary as well. The data imaging does not support a model of a broad depression of the seismic boundary. Nor is the boundary deepened by more than 40 km. If I have imaged the seismic discontinuity at 675 km in and around the slab, then this discontinuity must be temperature related and due to the spinel to perovskite phase transition. This does not mean there is no chemical change around that depth, only that the seismic data sense the phase transition there. A chemical boundary in this region would be fairly invisible to seismic data, and such a compositional change would not necessarily be associated with the 670 km depth anyway (Jeanloz, 1991). In contrast, all models of upper mantle bulk composition have mineral assemblages which undergo phase transitions at pressures consistent with the 670 km depth.

Another consideration is whether or not the temperature related discontinuity is actually due to the cold slab itself, or just a thermal anomaly projected into the lower mantle by a cold slab impinging upon, but not passing through, a chemical

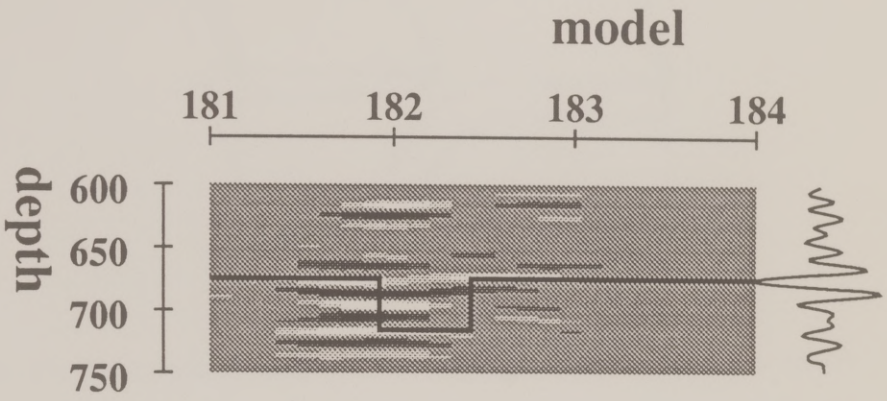
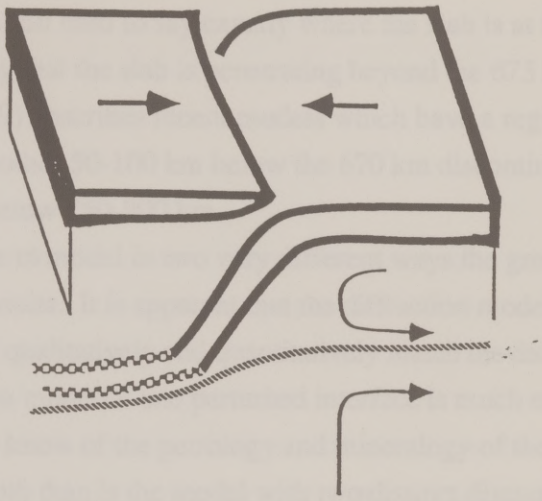


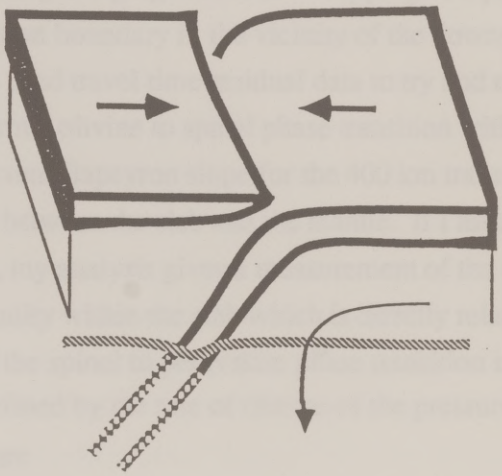
Figure 24) The data image of Figure 20 (b) with the trace of the best fit model for the 675 interface. The bright parts of the image correspond with the model position.

Figure 25) Compositional change (above) versus phase transition (below). In the former case, convection would be separate between the upper and lower mantle and the subducting slab would be deflected at the base of the upper mantle. Presumably any seismic boundary related to this would be broadly depressed over the region. A phase transition would not necessarily prevent the descending slab from penetrating into the lower mantle, and the seismic boundary would be confined in space relative to the position of the slab and the thermal structure.

Chemical boundary



Phase transition



boundary at 675 km. Presumably, this type of thermal anomaly would be broad, and, as I have said, my data analysis does not support this description of the seismic boundary. It is still hard to say exactly where the slab is at these depths, but the result does imply that the slab is penetrating beyond the 675 km depth in this region. Anderson (1989) describes recent models which have a region of high velocity gradient for another 50-100 km below the 670 km discontinuity, with the lower mantle starting below 750-800 km.

I have been able to model in two very different ways the gross aspects of the real data imaging results. It is apparent that the diffraction model produces synthetic images which qualitatively and quantitatively match the data better. Certainly, the diffraction model of one perturbed interface is much easier to reconcile with what we know of the petrology and mineralogy of the Earth at around 650-700 km depth than is the model with two distinct discontinuities in that region.

Thus, I interpret my imaging results as a mapping of a perturbed spinel to perovskite phase transition boundary in the vicinity of the downgoing slab. Solomon and U (1975) used travel time residual data to try and measure the elevation of the exothermic olivine to spinel phase transition within the slab. They use their result and a given Clapeyron slope for the 400 km transition to estimate the temperature difference between the slab and the mantle. If I assume a value for the temperature difference, my analysis gives a measurement of the 41 km depression of the 670 km discontinuity within the slab which is directly related to measuring the Clapeyron slope of the spinel to perovskite phase transition itself. The Clapeyron slope γ is defined by the rate of change of the pressure with respect to the change in temperature

$$\gamma \equiv \left(\frac{dp}{dT} \right) . \qquad (6)$$

The increase in pressure with changes in depth is given by the density ρ and the gravitational acceleration g

$$\frac{dp}{dz} = \rho g \ . \qquad (7)$$

From equations 6 and 7, I get an equation for the Clapeyron slope

$$\gamma = \rho g \left(\frac{\delta z}{\delta T} \right) \ . \qquad (8)$$

To evaluate equation 8, we need to know the acceleration of gravity, the density, a change in temperature and a change in depth for the spinel to perovskite transition. My inversion through forward modeling result gives a measurement of the change in depth, $\delta z = 41$ km, of the discontinuity between where it is in the mantle and where it is within the slab. We can take the acceleration of gravity to be a constant and equal 10 m/s^2. We can only estimate the density, but a very probable value (from PREM and other models) is 4000 kg/m^3. The change in temperature must also be estimated. With a value for the temperature difference $\delta T = -700$ K (Machetel and Weber, 1991), I calculate a Clapeyron slope of $\gamma = -2.3$ MPa/K. If I increase the temperature difference to around -900 K, the Clapeyron slope is reduced only a little, to $\gamma = -1.8$ MPa/K. With a reasonable value of $dT = -800$ K, the Clapeyron slope for the spinel to perovskite phase transition is $\gamma = -2.1$ MPa/K. These values are all very close to the commonly assumed values for the endothermic phase transition.

The importance of this measurement becomes clear when we consider recent convection calculations. Introducing the spinel to perovskite phase transition into a realistic mantle model, and using a preferred Clapeyron slope value of $\gamma = -2.0$ MPa/K from other work, Machetel and Weber (1991) found intermittent mixing between the upper and lower mantle in their convection calculations. My result supports the study of Machetel and Weber (1991) which may help to reconcile the seemingly opposing data and the differing views related to mantle convection. The results of a recent seismological study using travel time tomography to image the subducting slab is consistent with this same theory, finding evidence for slab penetration into the lower mantle in some places, and slab deflection at the upper-lower mantle boundary in others (van der Hilst, et al.).

Other S to P conversions

The data sampling of the 410 km discontinuity is poor for the SdP analysis, since most of the sources are deeper than that depth. In Figure 19 I can see an event at the right depth, and located correctly for where the data sample. The single trace stack tries to pull it out, but it is very weak compared to the spike from the 670 km discontinuity. This is primarily due to the lack of data sampling of this boundary.

Figure 19 also shows a possible discontinuity at around 520 km in the 2-D image. Variations in the sharpness of a 520 km discontinuity may indicate changes in chemical composition; however, depth changes of this transition are controlled mostly by thermal structure (Liu, 1980). Recent long period seismic data analyses (Shearer, 1990a,b; Revenaugh and Jordan; 1989; Revenaugh, 1990) support the existence of a discontinuity at a depth of around 520 km. My imaging hints at structure around 520 km depth, but any S to P conversion at that depth seems to be extremely weak. As with the '410 km' discontinuity, the data sampling of the 520 km depth is not as good as that of the 670 km depth; however, it is better than the 410 km depth, and the single trace stack does not bring out anything significant at 520 km.

Upper mantle structure and seismic reflections

An interesting result from my analysis is that I do not see coherent scattered energy in the P coda from underside p to P reflections or s to P conversions. Figures 26 and 27 show the imaging results from the delay-time moveout and stacking analyses using equations 3 and 4. I might expect to see at least the '410 km' discontinuity, and perhaps the LVZ.

Jeanloz and Thompson (1983) consider the possible phase transitions for the 400 and 670 km discontinuities. They determine that the transition at 400 km from olivine to spinel is a discontinuous (univariant) reaction which can occur sharply at mantle pressures and temperatures. Since seismically sharp boundaries are seen at those depths, it would seem that reaction kinetic effects are not

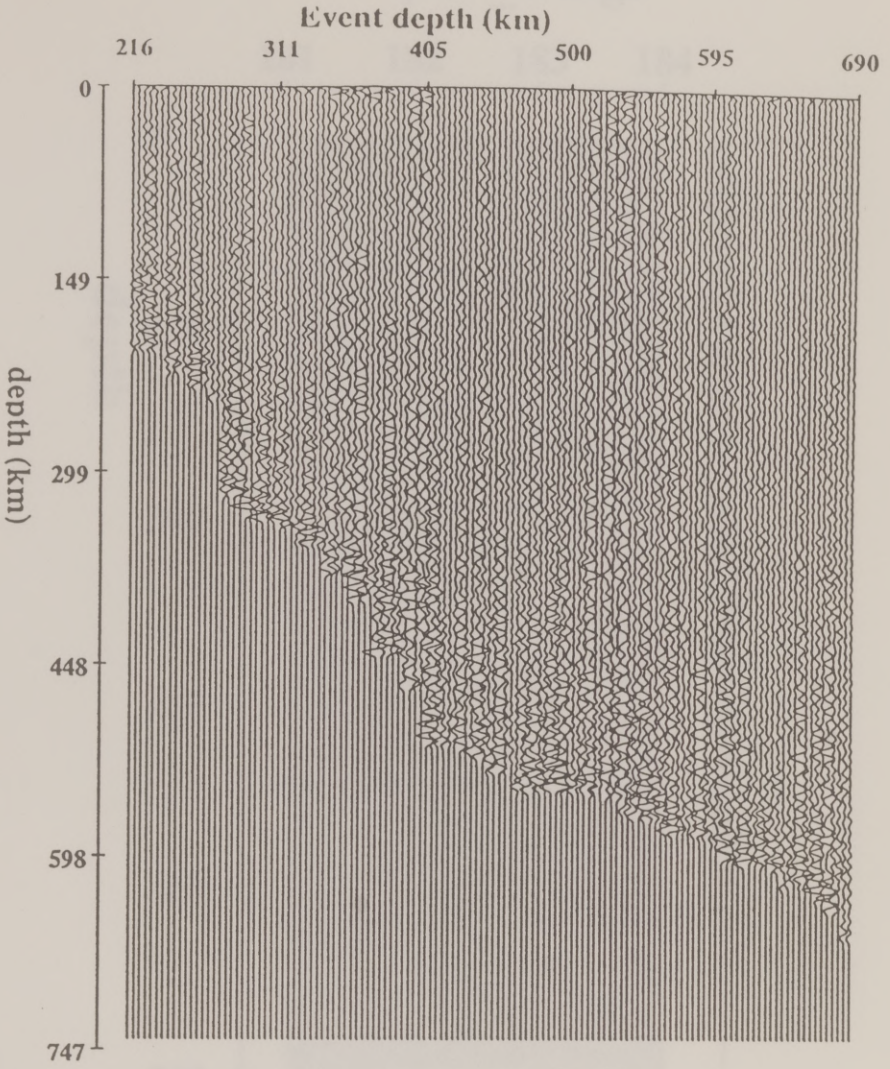


Figure 26 a) Depth section of data records after pdP analysis. Records plotted as a function of earthquake event depth. pP (seafloor reflection) is mapped to zero depth.

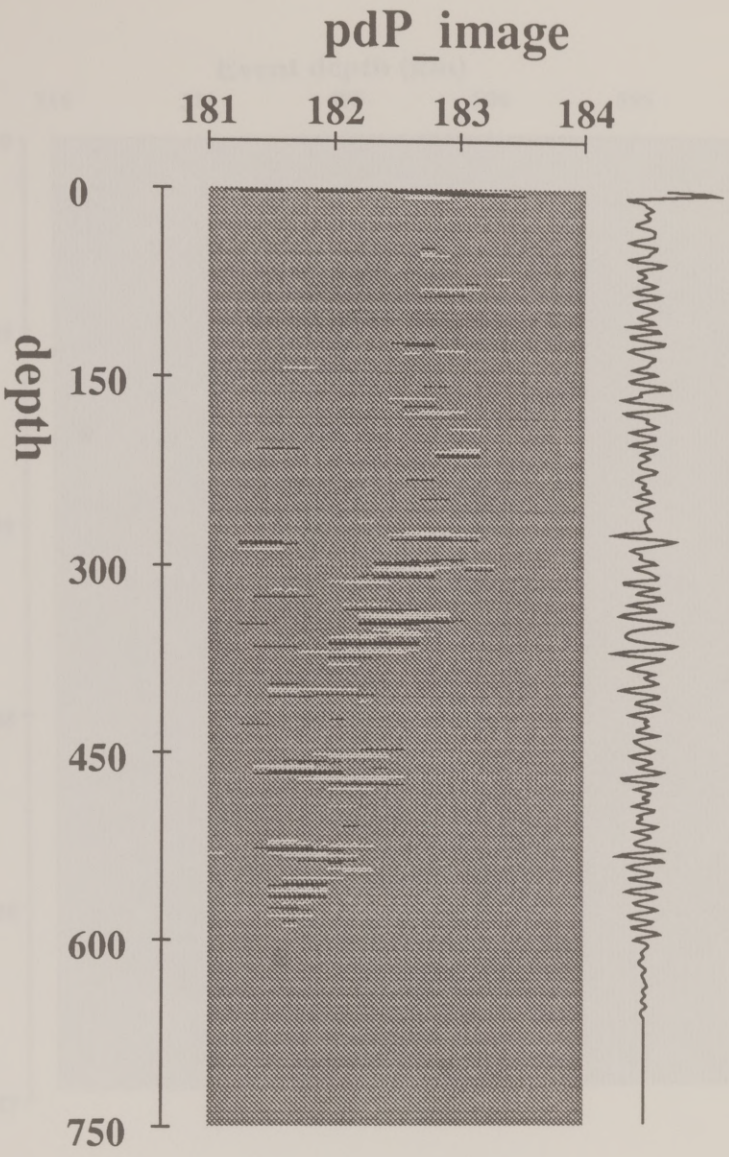


Figure 26 b) Depth records after pdP analysis stacked into image plane perpendicular to trench axis. Single trace stack on right for comparison. There is only a hint of P wave reflectivity structure at around 300 km, and little else in the image.

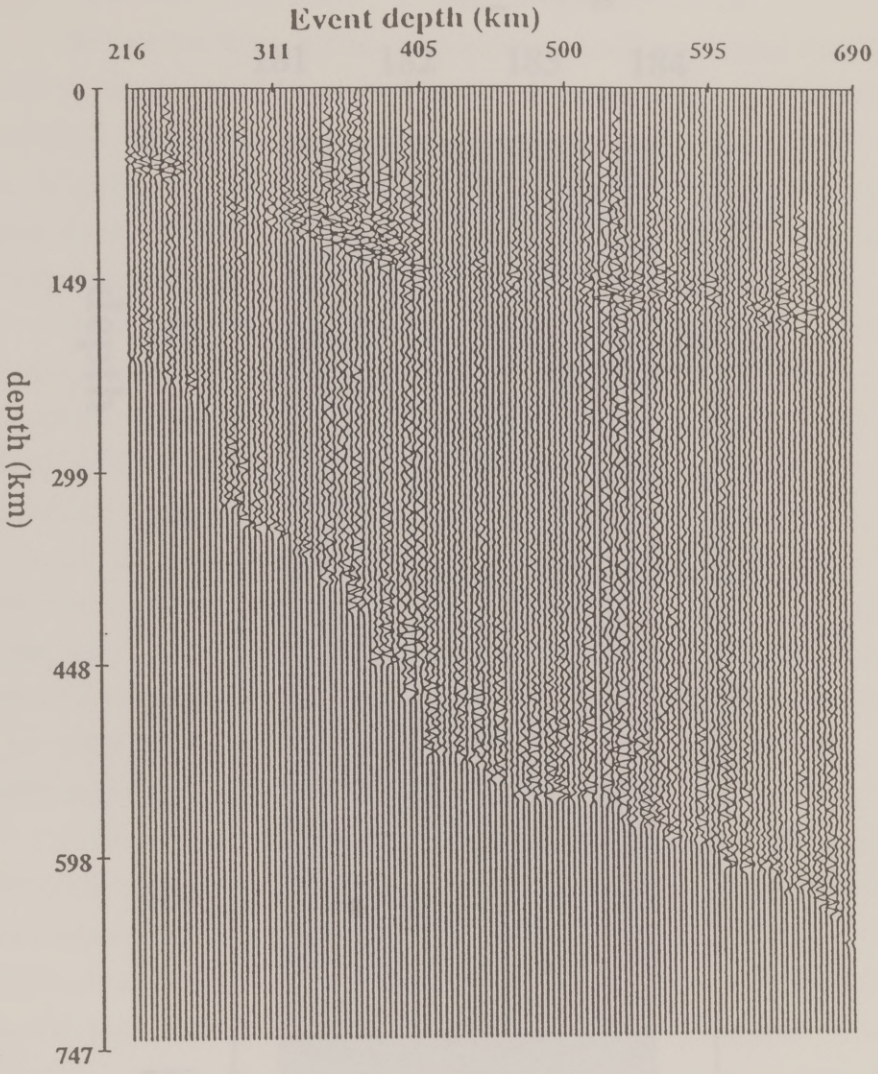


Figure 27 a) Depth section of data records after sdP analysis. Records plotted as a function of earthquake event depth. The pP phase is improperly moved out, ending up in the structure at around 100-150 km depth.

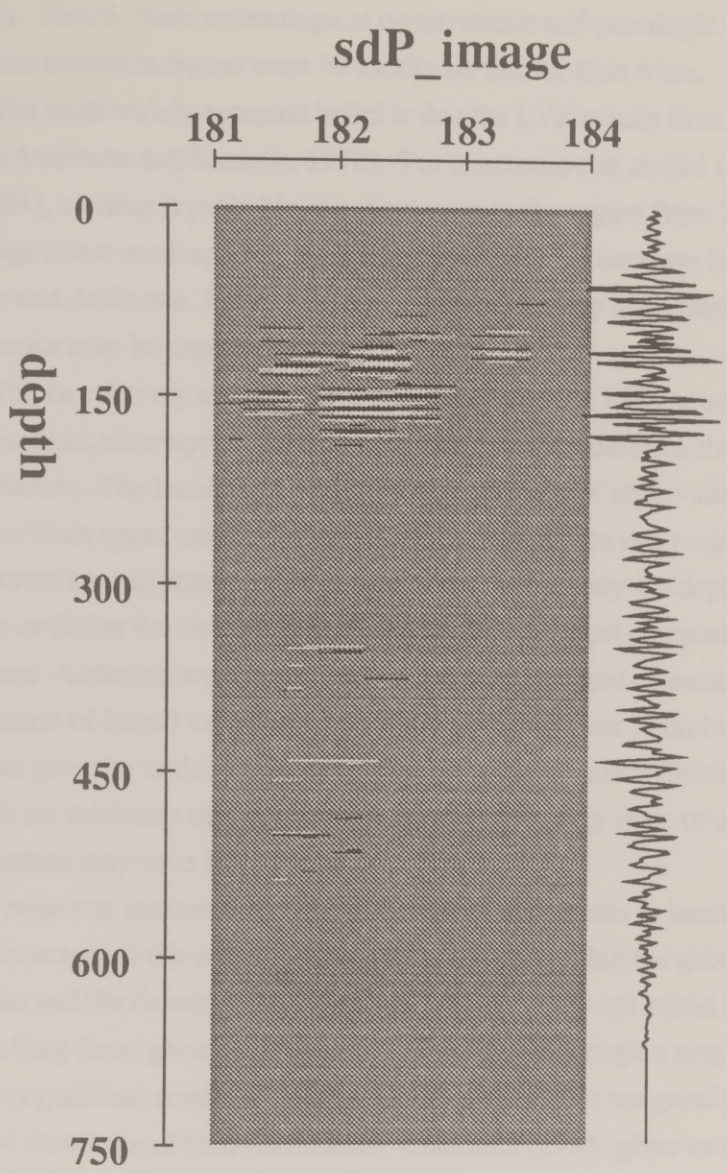


Figure 27 b) Depth records after sdP analysis stacked into image plane perpendicular to trench axis. Single trace stack on right for comparison. There is no apparent sdP structure that stacks out of the noise. The shallow features are due to the high amplitude pP phases as seen in Figure 27 (a).

important. Hence, from seismological observations and petrologic considerations, the 400 km transition region must be locally no thicker than 4 km.

The most widely accepted belief is that the LVZ results from a partial melt fraction (Anderson and Sammis, 1970). For a tectonic-rise model (Anderson and Bass, 1984), melting is probably at a maximum in the region from 75-200 km. Knowledge that a small amount of melt can cause a large decrease in velocity (Spetzler and Anderson, 1968) would suggest that the top and bottom of the low velocity zone may be sharp reflectors.

The fact that my analyses have failed to image the LVZ or underside reflections from other upper mantle structure can help to describe the possible discontinuities. The incoherency of scattering and lack of observations of reflections from upper mantle discontinuities can be due to two major factors. These factors are temperature differences which would vary the depth of a phase transition or define the amount of partial melt, and chemical composition differences. Anderson and Bass (1984) mention that crystal orientation may be another cause of lateral variations in velocity above 400 km. Combinations of these causes can give rise to three effects: a) overall velocities are different from region to region, b) the thickness of a transitional discontinuity may vary, or c) the discontinuities may have lateral structural variations.

I expect to encounter the first effect (a) of differences in laterally averaged velocity structures from one subduction zone to another, but the geometry of acquisition and the density of the receivers and sources is not suited to distinguishing these gross differences. Transmission through a single subduction zone with significant small scale lateral heterogeneity does not greatly affect the first arrivals of forward scattered wavefields. Claerbout (1985) gives an example of the 'healing' of an irregular wavefront, which can simulate a disrupted wavefront due to lateral heterogeneity. A coda develops behind the nearly uniform wavefront. Of course, for large scale heterogeneities, such as the slab, there may be large effects, but these are coherent across a wide frequency range. Silver and Chan (1985) used this idea in their study of P codas and slab penetration.

Considering effect (b), Lees et al. (1983) consider the P'670P' phase to determine the reflection properties of the 670 km discontinuity. They calculate

frequency dependent reflection coefficients and showed how high frequencies are substantially reduced for a transitional boundary, as opposed to a first order discontinuity model. Ward (1978) has shown that a transition zone of 10 km is still a sharp interface to long period waves, although higher frequencies (periods<3 sec) in the reflections are negligible. Calculating synthetic seismograms, Davis et al. (1989) determine that their null P'P' precursor results are consistent with either cutting the impedance jump for the 670 km discontinuity of PREM in half, smoothing the impedance jump over a 15-30 km thick transition zone, or introducing small amplitude (10 km) topography on the discontinuity with a lateral wavelength of 300 km.

Since the underside reflection imaging of Figures 26 and 27 fail to bring out even the '400 km' discontinuity, it is very likely that topography of the discontinuity has worked to defocus the reflected energy. As in Figure 13 where the free surface image was messed up due to the bathymetry. The data coverage as we get farther away from the earthquake hypocenters is probably not sufficient to resolve the scattering due to even a modest amount of topography. In the subduction zone environment, it is easy to imagine quite a bit of lateral heterogeneity as a direct consequence of reactions due to the cold lithospheric slab going down into the hot mantle. Ringwood (1975) describes a dynamic petrologic model where the mantle overlying the Wadati-Benioff zones between 100-300 km depths has been mobilized by the sinking slab, and rising diapirs of mantle pyrolite produce basalts as pressure is released. Anderson (1987) notes that the existence of the thick, deep LVZ in tectonic provinces gives an alternative to the hypothesis for a shallow pyrolite-like reservoir for basaltic magmas. Convective patterns, diapirism and the perturbations of the thermal structure resulting from either theory are plausible causes of lateral variation in upper mantle reflectors.

We might postulate three types of lateral variation of boundaries in the upper mantle. Applying a Kirchhoff diffraction calculation for a flat but discontinuous reflector shows a loss of both high and low frequencies as we depart from a continuous boundary. For a single scatterer, the character of both the high and the low frequency reflected (scattered) pulses are similar. As the scatterer is reduced in size, both reflections lose amplitude and the pulse shapes change because the

integration is over a smaller fraction of Fresnel zones. It is clear that a very slight amount topography can substantially reduce the high frequencies in a reflected pulse, yet allow longer period information to be scattered (Davis et al., 1989), and this can be understood easily in terms of Kirchhoff diffraction theory. When considering existing discontinuities, I expect short period seismic observations to be most dependent on lateral structural variation of the reflectors. My analyses cannot distinguish between the non-existence of boundaries such as at 200 km over the possibility that topography of the boundary or a transitional boundary has defocussed the back-scattered energy.

CONCLUSIONS

As in determining the gross features of the interior of the Earth, seismology can be the leading geophysical technique to help map the deviations from smooth models. This requires researchers to study specialized data sets which have sampled specific regions of the Earth. The imaging method I have applied is a general one, using ideas from inverse theory and seismic data processing techniques, and it can be made more elaborate as required by a more elaborate data set. Certainly, the first extension of this research would be to get more data, and then try and solve the source problem more explicitly as well. With modern broad-band data collection and the trend with global structure workers moving away from strictly long-period seismological studies, earth scientists have a better chance of characterizing the local variations from the globally averaged velocity models. In applying my imaging and inversion method to S to P scattered energy in the P wave coda of records from deep earthquakes in Tonga, I was able to measure the depth difference of the spinel to perovskite phase transition between the mantle and the slab. Using my measurement of a 41 km depth difference, I calculated a Clapeyron slope value of γ=-2.1 MPa/K for the transition which agrees with other experimental and theoretical values. Machetel and Weber (1991) used this value for the Clapeyron slope in their convection calculations to produce intermittent mixing between the upper and lower mantle. My result supports a description of the 670 km discontinuity in terms of an endothermic phase change from spinel to perovskite, and this can help reconcile the seismic observations. The presence of

the phase transition disrupts full mantle convection patterns, allowing only intermittent mixing in an otherwise chemically homogeneous mantle.

APPENDIX I Some Scattering Theory

General scattering theory analysis comes out of the study of perturbations to a known wave equation. We start with a homogeneous (source-less) wave equation given by the differential operator ζ_o,

$$\zeta_o(\psi) = 0 \ . \tag{1}$$

We introduce a model perturbation to give a new differential operator ζ,

$$\zeta(\psi(x)) = (\zeta_o - \delta m(x)) \psi(x) = 0 \ . \tag{2}$$

This describes scattering due to a potential (δm) and the result is a forcing term (source) in the original wave equation

$$\zeta_o(\psi(x)) = \delta m(x) \psi(x) \ . \tag{3}$$

We let ψ_i be the solution to the homogeneous equation, and ψ_s be the scattered wavefield. Then

$$\psi(x) = \psi_i + \psi_s \ . \tag{4}$$

Using equation 4 in equation 3, we get

$$\zeta_o (\psi_i + \psi_s) = \zeta_o (\psi_s) = \delta m \psi \ . \tag{5}$$

Given the Green's function solution Ψ_o which satisfies, for a delta function source at x_s,

$$\zeta_o \left(\Psi_o(x,x_s) \right) = \delta(x-x_s) \ , \tag{6}$$

we can solve for the scattered wavefield as

$$\psi_s = \int_V \Psi_o(x,x') \ \delta m(x') \ \psi(x') \ d^3x' \ . \tag{7}$$

Thus, the wavefield ψ at some observation point x_o is given by

$$\psi(x_o) = \psi_i(x_o) + \int_V dv \ \Psi_o(x_o,x') \ \delta m(x') \ \psi(x') \ . \tag{8}$$

If the volume integral of equation 7 is restricted to a surface (i.e., the perturbation δm is zero except on a given boundary), this result can be related to the Kirchhoff integral of diffraction theory. Here, we have the scattered wavefield given by the values of the wavefunction within the volume. The normal method for solving equation 8 is by iteration, starting with $\psi(x')$ equal to $\psi_i(x')$ within the integral for the scattered wavefield. If we stop after this first iteration, we get the Born approximation to the Neumann series solution for ψ. Given a source at x_s, $\psi_i(x) = \Psi_o(x,x_s)$. Then

$$\psi(x_o) = \Psi_o(x_o,x_s) + \int_V dv \ \Psi_o(x_o,x') \ \delta m(x') \ \Psi_o(x',x_s) \ . \tag{9}$$

Equation 9 gives the linearized solution to the volume scattering problem, and the result is intuitive. $\Psi_o(x',x_s)$ gives the propagation of the wave from the source point to the scattering point x', and $\Psi_o(x_o,x')$ gives the propagation from the scattering point to the observation point. The weight given the scattering contributions is given by the perturbation δm. Starting with a summation over all possible scattering paths in equation 8, the linearized solution in equation 9 implies only single scattering is important, with all multiple scattering neglected. We can associate equation 9 with a Taylor series, where the integral defines the first derivative of the wavefield with respect to the model perturbation. This is called the Frechet derivative of $\psi(x)$, and the operator of equation 9 would have a Frechet kernel, or sensitivity function given by $\Psi_o(x_o,x')\Psi_o(x',x_s)$.

APPENDIX II A Bit of Diffraction Theory

Diffraction theory comes out of the study of media obstructed by a surface or boundary. In most boundary value problems, the boundary conditions are normally specified in terms of the wave amplitude and its normal derivative at the boundary. For two solutions to the wave equation, ψ_1 and ψ_2, let us consider the following integral

$$\int_S \psi_1 \left(\hat{\mathbf{n}} \bullet \nabla \psi_2 \right) ds \ . \tag{1}$$

This analysis will start with Gauss' Theorem which states, for vector **A**,

$$\int_{\Delta V} \nabla \bullet \mathbf{A} \, dV = \int_S \mathbf{A} \bullet ds \ . \tag{2}$$

Rearranging equation 1, and then applying Gauss' Theorem, we obtain

$$\int_S \psi_1 \left(\hat{\mathbf{n}} \bullet \nabla \psi_2 \right) ds = \int_S \left(\psi_1 \nabla \psi_2 \right) \bullet \hat{\mathbf{n}} \, ds = \int_V \nabla \bullet \left(\psi_1 \nabla \psi_2 \right) dv \ . \tag{3}$$

Now,

$$\nabla \bullet \left(\psi_1 \nabla \psi_2 \right) = \nabla \psi_1 \bullet \nabla \psi_2 + \psi_1 \nabla^2 \psi_2 \ . \tag{4}$$

Interchanging the wavefunctions in equation 3, using equation 4, and subtracting gives a result known as Green's Theorem

$$\int_S \left(\psi_1 \nabla\psi_2 - \psi_2 \nabla\psi_1 \right) \bullet ds = \int_V \left(\psi_1 \nabla^2\psi_2 - \psi_2 \nabla^2\psi_1 \right) dv \ . \qquad (5)$$

Since ψ_1 and ψ_2 satisfy the wave equation, we have that

$$\nabla^2\psi_i = -k^2\psi_i \ , \text{ for } i = 1,2 \ , \qquad (6)$$

and so the right hand side of equation 5 is

$$\int_V \left(\psi_1(-k^2\psi_2) - \psi_2(-k^2\psi_1) \right) dv = 0 \ . \qquad (7)$$

Now, the left hand side of equation 5 is an integral over a closed surface excluding the field point x_o of interest but including the volume in the integral of equations 3-7. Taking ψ_1 to be the wavefunction ψ we are interested in evaluating, and ψ_2 as the Green's function solution Ψ_o to the wave equation, the surface integration of equation 5 reduces to two integrals: an integral which collapses to the point xo, and yields the wave amplitude at that point, and an integral over the rest of the surface given by ΔS

$$\psi(x_o) \propto \int_{\Delta S} \left(\psi \nabla\Psi_o - \Psi_o \nabla\psi \right) \bullet ds \ . \qquad (8)$$

Thus, the value of ψ at the observation point is dependent upon the values it takes on the surface of integration. This is a surface scattering equation from diffraction

theory, and is the basis for the Kirchhoff integral for forward scattering solutions as well as for Kirchhoff migration algorithms.

APPENDIX III Calculus-based Inverse Theory

We can begin the inverse analysis by writing a Taylor series for a function f(**x**) about a point $\mathbf{x}_o$ from the origin as

$$f(\mathbf{x}) = f(\mathbf{x}_o) + \sum_i \frac{\partial f}{\partial x_i} x_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} x_i x_j + \ldots , \tag{1}$$

where the vector of independent variables $\mathbf{x}_o = (x_1, x_2, x_3, \ldots x_n)$. If we drop the terms of order x^3 and higher, we get a relation of the form

$$f(\mathbf{x}) \cong C - \mathbf{b} \bullet \mathbf{x} + \frac{1}{2} \mathbf{x} \bullet \mathbf{A} \bullet \mathbf{x} , \tag{2}$$

where C is a constant, **b** is the direction of steepest descent, or the negative of the gradient of the function f at point $\mathbf{x}_o$, and **A** is the matrix of second derivatives, or the Hessian of the function f at point $\mathbf{x}_o$, given by

$$C \equiv f(\mathbf{x}_o) , \quad \mathbf{b} \equiv -\nabla f |_{x_o} , \quad \text{and} \quad [\, \mathbf{A} \,]_{ij} \equiv \frac{\partial^2 f}{\partial x_i \partial x_j} |_{x_o} . \tag{3}$$

From equation 2, we can get the gradient of the function at point **x**

$$\nabla f(\mathbf{x}) = -\mathbf{b} + \mathbf{A} \bullet \mathbf{x} . \tag{4}$$

To find the minimum (or an extremum) of the function f(**x**), we require that the gradient vanish. Thus, at the minimum point $\mathbf{x}_m$,

$$\mathbf{A} \bullet \mathbf{x}_m = \mathbf{b} \ . \tag{5}$$

We get the vector which minimizes the function f(**x**) from equation 5 by inverting the Hessian matrix **A**

$$\mathbf{x}_m = \mathbf{A}^{-1} \mathbf{b} \ . \tag{6}$$

For the seismic inverse problem, we start with a least-squares functional over model space given by

$$F(\mathbf{m}) = \frac{1}{2} \left[(\delta \mathbf{u}_o - \mathbf{G}\delta\mathbf{m})^t \mathbf{C}_u^{-1} (\delta \mathbf{u}_o - \mathbf{G}\delta\mathbf{m}) + \delta\mathbf{m}^t \mathbf{C}_m^{-1} \delta\mathbf{m} \right] , \tag{7}$$

where **G** is the linear operator, or Frechet derivative for the forward problem describing the relationship between the data and the model $\mathbf{u}=g(\mathbf{m})$. We wish to find the minimum of the functional F(**m**) by jointly minimizing the data error ($\delta\mathbf{u}_o$-$\mathbf{G}\delta\mathbf{m}$) and the distance from some reference model $\delta\mathbf{m}=(\mathbf{m}_o-\mathbf{m})$. As earlier, we find the gradient of the functional F with respect to model parameters **m**, and set it equal to zero to find the minimum

$$\nabla F = \mathbf{G}^t \mathbf{C}_u^{-1} (\delta \mathbf{u}_o - \mathbf{G}\delta\mathbf{m}) - \mathbf{C}_m^{-1} \delta\mathbf{m} = 0 \ . \tag{8}$$

Rearranging the terms, we find

$$\left(\mathbf{C}_m \mathbf{G}^t \mathbf{C}_u^{-1} \mathbf{G} + \mathbf{I} \right) \delta\mathbf{m} = \mathbf{C}_m \mathbf{G}^t \mathbf{C}_u^{-1} \delta\mathbf{u}_o \ . \tag{9}$$

Comparing this with equation 5, we see that the steepest descent direction at point $\mathbf{m}_o$ is given by the right hand side of equation 9, while the Hessian at $\mathbf{m}_o$ is given by the factor premultiplying $\delta\mathbf{m}$ on the left hand side of that equation. We can invert the Hessian to find the solution $\delta\mathbf{m}$ as in equation 6. From equation 9 it is clear that the seismic inverse problem is very much dependent upon the linearization $\mathbf{G}$ of the forward problem. When the problem is nonlinear, i.e. the data is a nonlinear function of the model parameters, equation 9 is solved a number of times in an iterative analysis. It is hoped that after a few iterations, the analysis will converge to the global minimum of $F(\mathbf{m})$.

REFERENCES

Adams, R. D., 1968, Early reflections of P'P' as an indication of upper mantle structure: Bull. Seis. Soc. Am., **58**, 1933-1947.

Aki, K., 1969, Analysis of the seismic coda of local earthquakes as scattered waves: J. Geophys. Res., **74**, 615-631.

Aki, K. and Richards, P. G., 1980, Quantitative seismology: theory and methods, **Vol. I**: W. H. Freeman and Company.

Anderson, D. L., 1967, Phase changes in the upper mantle: Science, **157**, 1165-1173.

_____, 1987, The depths of mantle reservoirs: in Magmatic processes: physiochemical principles, edited by B. O. Mysen, The Geochemical Society, Special Publication No. 1.

_____, 1989, Theory of the Earth: Blackwell Scientific Publications.

Anderson, D. L. and Sammis, C., 1970, Partial melting in the upper mantle: Phys. Earth Planet. Inter., **3**, 41-50.

Anderson, D. L. and Bass J. D., 1984, Mineralogy and composition of the upper mantle: Geophys. Res. Lett., **11**, 637-640.

Backus, G. and Gilbert, F., 1968, The resolving power of gross earth data: Geophys. J. R. astr. Soc., **16**, 169-205.

_____, 1970, Uniqueness in the inversion of inaccurate gross earth data: Phil. Trans. Roy. Soc. London, **266a**, 123-192.

Barazangi, M., Pennington, W. and Isacks, B., 1975, Global study of seismic wave attenuation in the upper mantle behind island arcs using pP waves: J. Geophys. Res., **80**, 1079-1092.

Barley, B. J., Hudson, J. A. and Douglas., A., 1982, S to P scattering at the 650 km discontinuity: Geophys. J. R. astr. Soc., **69**, 159-172.

Bock, G. and Kind, R., 1991, A global study of S-to-P and P-to-S conversions from the upper mantle transition zone: Geophys. J. Int., **107**, 117-129.

Bolt, B. A., O'Neill, M. and Qamar, A., 1969, Seismic waves near 110°: Is structure in core or upper mantle responsible?: Geophys. J. R. astr. Soc., **16**, 475-487.

Bott, M. H. P., 1982, The interior of the Earth: Elsevier Science Pub. Co., Inc.

Bortfeld, R., 1989, Geometrical ray theory, rays and traveltimes in seismic systems (second-order approximations of the traveltimes): Geophysics, **54**, 342-349.

Bregman, N. D., Bailey, R. C. and Chapman, C. H., 1989, Crosshole seismic tomography: Geophysics, **54**, 200-215.

Cardimona, S., 1991, Waveform inversion and digital filter theory: Geophysics, **56**, 534-536.

Cardimona, S. and Garmany, J., 1989, Local linearization and nonuniqueness of Frechet derivative kernels: Transactions of the American Geophysical Union, EOS, **70**, 1221-1222.

_____, 1991, How rays feel about themselves: in preparation for Geophysics.

Carrion, P., 1991, Dual tomography for imaging complex structures, Geophysics, **56**,1395-1404.

Cerveny, V. and Hron, F., 1980, The ray series method and dynamic ray tracing system for three-dimensional inhomogeneous media: Bull. Seism. Soc. Am., **70**, 47-77.

Claerbout, J. F., 1985, Fundamentals of geophysical data processing: Blackwell Scientific Publications.

Crase, E., Pica, A., Noble, M., McDonald, J. and Tarantola, A., 1990, Robust elastic nonlinear waveform inversion, application to real data: Geophysics, **55**, 527-538.

Creager, K. C. and Jordan, T. H., 1986, Slab penetration into the lower mantle beneath the mariana and other island arcs of the northwest pacific: J. Geophys. Res., **91**, 3573-3589.

Davis, J. P., Kind, R. and Sacks, I. S., 1989, Precursors to P'P' re-examined using broad-band data: Geophys. J. Int., **99**, 595-604.

Devaney, A. J. 1984, Geophysical diffraction tomography: IEEE Transactions on Geoscience and Remote Sensing, **GE-22**, 3-13.

Docherty, P., 1991, A brief comparison of some Kirchhoff integral formulas for migration and inversion: Geophysics, **56**, 1164-1169.

Dziewonski, A. M. and Anderson, D. L., 1981, Preliminary reference earth model: Phys. Earth Planet. Inter., **25**, 297-356.

Elmore, W. C. and Heald, M. A., 1969, Physics of waves: Dover Publications Inc.

Engdahl, E. R. and Flinn, E. A.,1969, Seismic waves reflected from discontinuities within the earth's upper mantle: Science, **163,** 177-179.

Esmersoy, C. and Miller, D., 1989, Backprojection versus backpropagation in multidimensional linearized inversion: Geophysics, **54**, 921-926.

Feynman, R. P. and Hibbs, A. R., 1965, Quantum mechanics and path integrals: McGraw-Hill Inc.

Frohlich, C., 1989, The nature of deep-focus earthquakes: Ann. Rev. Earth Planet. Sci., **17**, 227-254.

Giardini, D. and Woodhouse, J. H., 1984, Deep seismicity and modes of deformation in Tonga subduction zone: Nature, **307**, 505-509.

Goupillaud, P. L., 1961, An approach to inverse filtering of near-surface effects from seismic records: Geophysics, **26**, 754-760.

Grand, S. P. and Helmberger, D. V., 1984a, Upper mantle shear structure of N. America: Geophys. J. R. astr. Soc., **76**, 399-438.

_____, 1984b, Upper mantle shear structure beneath the northwest Atlantic ocean: J. Geophys. Res., **89**, 11465-11475.

Grand, S. P. and Ding, X. Y., 1989, Residual spheres and slab penetration into the lower mantle: Transactions of the American Geophysical Union, EOS, **70**, 1322.

Graves, R. W. and Clayton, R. W., 1990, Modeling acoustic waves with paraxial extrapolators: Geophysics, **55**, 306-319.

Haddon, R. A. W., Husebye, E. S. and King, D. W., 1977, Origins of precursors to P'P': Phys. Earth Planet. Inter., **14**, 41-70.

Hager, B. H., 1984, Subducted slabs and the geoid: constraints on mantle rheology and flow: J. Geophys. Res., **89**, 6003-6015.

Hales, A. L., Muirhead, K. J. and Rynn, J. M. W., 1980, A compressional velocity distribution for the upper mantle: Tectonophysics, **63**, 309-348.

Herraiz, M. and Espinosa, A. F., 1987, Coda waves: a review: PAGEOPH, **125,** 499-577.

Hudson, J. A. and Heritage, J. R., 1981, The use of the Born approximation in seismic scattering problems: Geophys. J. R. astr. Soc., **66**, 221-240.

Jeanloz, R., 1991, Effects of phase transitions and possible compositional changes on the seismological structure near 650 km depth: Geophys. Res. Lett., **18**, 1743-1746.

Jeanloz, R. and Thompson, A. B., 1983, Phase transitions and mantle discontinuities: Rev. Geophys. Space Phys., **21**, 51-74.

Kennett, B. L. N., 1984, An operator approach to forward modelling, data processing and migration: Geophys. Prosp., **32**, 1074-1090.

King, D. W., Husebye, E. S. and Haddon, R. A. W., 1976, Processing of seismic precursor data: Phys. Earth Planet. Inter., **12**, 128-134.

Kolb, P. and Canadas, G., 1986, Least-squares inversion of prestack data, simultaneous identification of density and velocity: Presented at the 16th Conf. on math geophys.

Lees, A. C., Bukowinski, M. S. T. and Jeanloz, R., 1983, Reflection properties of the phase transition and compositional change models of the 670-km discontinuity: J. Geophys. Res., **88**, 8145-8159.

Lines, L. R., and Treitel, S. 1984, Tutorial: A review of least-squares inversion and its application to geophysical problems: Geophys. Prosp., **32**, 159-186.

Lines, L. R., Schultz, A. K. and Treitel, S., 1988, Cooperative inversion of geophysical data: Geophysics, **53**, 8-20.

Lines, L. R. and La Fehr, E. D., 1989, Tomographic modeling of a cross-borehole data set: Geophysics, **54**, 1249-1257.

Liu, L-G, 1980, The pyroxene-garnet transformation and its implication for the 200-km seismic discontinuity: Phys. Earth Planet. Inter., **23**, 286-291.

Luo, Y. and Schuster, G. T., 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645-653.

McAulay, A. D., 1986, Plane-layer prestack inversion in the presence of surface reverberation: Geophysics, **51**, 1789-1803.

McGillivray, P. R. and Oldenburg, D. W., 1990, Methods for calculating Frechet derivatives and sensitivities for the non-linear inverse problem, a comparitive study: Geophys. Prosp., **38**, 499-524.

McMechan, G. A. and Fuis, G. S., 1987, Ray equation migration of wide-angle reflections from southern Alaska: J. Geophys. Res., **92**, 407-420.

Machetel, P. and Weber, P., 1991, Intermittent layered convection in a model mantle with an endothermic phase change at 670 km: Nature, **350**, 55-57.

Mathews, J. and Walker, R. L., 1970, Mathematical methods of physics: Benjamin/Cummings Publishing Co.

Menke, W. 1984, Geophysical data analysis, discrete inverse theory: Academic Press Inc.

Michelena, R. J. and Harris, J. M., 1991, Tomographic traveltime inversion using natural pixels: Geophysics, **56**, 635-644.

Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52**, 1211-1228.

Mora, P., 1989, Inversion = migration + tomography: Geophysics, **54**, 1575-1586.

Moser, T. J., 1991, Shortest path calculation of seismic rays: Geophysics, **56**, 59-67.

Nataf, H.-C., Nakanishi, I. and Anderson, D. L., 1986, Measurements of mantle wave velocities and inversion for lateral heterogeneities and anisotropy: 3. Inversion: J. Geophys. Res., **91**, 7261-7307.

Oldenburg, D. W., 1981, A comprehensive solution to the linear deconvolution problem: Geophys. J. R. astr. Soc., **65**, 331-358.

Parker, R. L. 1977, Understanding inverse theory: Annual Review of the Earth and Planetary Sciences, **5**, 35-64.

Peacock, K. L. and Treitel, S., 1969, Predictive deconvolution: theory and practice: Geophysics, **34**, 155-169.

Pennington, W. D. and Isacks, B. L., 1979, Analysis of short-period waveforms of P phases from deep-focus earthquakes beneath the Fiji Islands: Geophys. J. R. astr. Soc., **56**, 19-40.

Pratt, R. G. and Goulty, N. R., 1991, Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data: Geophysics, **56**, 208-224.

Rajan, S. D. and Frisk, G. V. 1989, A comparison between the Born and Rytov approximations for the inverse backscattering problem: Geophysics, **54**, 864-871.

Revenaugh, J., 1990, Reflectivity of the transition zone: constraints on the 520-km discontinuity from ScS reverberations: Transactions of the American Geophysical Union, EOS, **71**, 1472.

Revenaugh, J., and Jordan, T. H., 1987, Observations of first-order mantle reverberations: Bull. Seis. Soc. Am., **77**, 1704-1717.

_____, 1989, A study of mantle layering beneath the western pacific: J. Geophys. Res., **94**, 5787-5813.

Richards, M. A. and Wicks, C. W., Jr., 1990, S-P conversion from the transition zone beneath Tonga and the nature of the 670 km discontinuity: Geophys. J. Int., **101**, 1-35.

Ringwood, A. E., 1975, Composition and petrology of the Earth's mantle: McGraw-Hill.

Schneider, W. A., 1978, Integral formulation for migration in two and three dimensions, Geophysics, **43**, 49-76.

Shearer, P., 1990a, Seismic imaging of upper-mantle structure with new evidence for a 520-km discontinuity, Nature, **344**, 121-126.

_____, 1990b, Long-period observations of the 520-km discontinuity: Transactions of the American Geophysical Union, EOS, **71**, 1472.

Silver, P. G. and Chan, W. W., 1985, Subducted slabs in the lower mantle: evidence from body wave multipathing: Transactions of the American Geophysical Union, EOS, **66**, 957-958.

Singh, R. P. and Singh, Y. P., 1991, RAYPT - A new inversion technique for geotomographic data: Geophysics, **56**, 1215-1227.

Solomon, S. C. and K. T. P. U, 1975, Elevation of the olivine-spinel transition in subducted lithosphere: seismic evidence: Phys. Earth Planet. Inter., **11**, 97-108.

Spetzler, H. and Anderson, D. L., 1968, The effect of temperature and partial melting on velocity and attenuation in a simple binary system: J. Geophys. Res., **73**, 6051-6060.

Squires, L. J. and Cambois, G., 1991, A linear filter interpretation of the inverse model covariance matrix: submitted to Geophysics.

Stark, P. B. and Frohlich, C., 1985, The depths of the deepest deep earthquakes: J. Geophys. Res., **90**, 1859-1869.

Stork, C. and Clayton, R. W., 1991, Linear aspects of tomography velocity analysis: Geophysics, **56**, 483-495.

Tarantola, A., 1984, Linearized inversion of seismic reflection data: Geophys. Prosp., **32**, 998-1015.

_____, 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893-1903.

_____, 1987, Inverse problem theory, methods for data fitting and model parameter estimation: Elsevier Science Publishing Company Inc.

Tarantola, A. and Valette, B., 1982, Generalized nonlinear inverse problems solved using the least-squares criterion: Rev. of Geophys. and Space Physics, **20**, 219-232.

Thorson, J. 1984, Velocity stack and slant stack inversion methods: Ph.D. thesis, Stanford University.

Tinti, S. and Ugolini, S., 1990, Pre-selection of seismic rays as a possible method to improve the inverse problem: Geophys. J. Int, **102**, 45-61.

Treitel, S. and Lines, L. R., 1982, Linear inverse theory and deconvolution: Geophysics, **47**, 1153-1159.

Treitel, S. and Robinson, E. A., 1969, Optimum digital filters for signal to noise ratio enhancement: Geophys. Prosp., **12**, 248-293.

Turcotte, D. T. and Schubert, G., 1982, Geodynamics: applications of continuum physics to geologic problems: John Wiley & Sons.

Turin, G. L. 1957, On the estimation in the presence of noise of the impulse response of a random, linear filter: IRE Transactions on Information Theory, 5-10.

van der Hilst, R., Engdahl, R., Spakman, W. and Nolet, G., 1991, Tomographic imaging of subducted lithosphere below northwest Pacific island arcs: Nature, **353**, 37-43.

van Trier, J. and Symes, W. W., 1991, Upwind finite-difference calculation of travel times: Geophysics, **56**, 812-821.

Van Wykhouse, R., 1973, SYNBAPS (synthetic bathymetric profiling systems): Tech. Rep. **TR-233**, Nav. Oceanogr. Office.

Vidale, J., 1988, Finite difference calculation of traveltimes: Bull. Seis. Soc. Am., **78**, 2062-2076.

_____, 1990, Finite-difference calculation of traveltimes in three dimensions: Geophysics, **55**, 521-526.

Vidale, J. and Ammon, C., 1989, Efficient seismic traveltime and amplitude calculations and applications to velocity inversion and migration: Transactions. of the American Geophysical Union, EOS, **70**, 1222.

Wapenaar, C. P. A., 1990, Representations of seismic sources in the one-way wave equations: Geophysics, **55**, 786-790.

Ward, S. N., 1978, Long-period reflected and converted upper-mantle phases: Bull. Seis. Soc. Am., **68**, 133-153.

Weber, M., 1990, Subduction zones--their influence on traveltimes and amplitudes of P-waves: Geophys. J. Int., **101**, 529-544.

Whitcomb, J. H. and Anderson, D. L., 1970, Reflection of P'P' seismic waves from discontinuities in the mantle: J. Geophys. Res., **75** , 5713-5728.

Wicks, C. W. and Richards, M. A., 1991, Effects of source radiation patterns on the phase S670P beneath the Tonga subduction zone: Geophys. J. Int., **107**, 279-290.

Wiggins, R. A., 1972, The general linear inverse problem: implications of surface waves and free oscillations for earth structure: Rev. of Geophys. and Space Physics, **10**, 251-285.

Williamson, P. R., 1990, Tomographic inversion in reflection seismology: Geophys. J. Int., **100**, 255-274.

_____, 1991, A guide to the limits of resolution imposed by scattering in ray tomography: Geophysics, **56**, 202-207.

Woodhouse, J. H. and Dziewonski, A. M., 1984, Mapping the upper mantle: three-dimensional modeling of earth structure by inversion of seismic waveforms: J. Geophys. Res., **89**, 5953-5986.

Xie, J. and Mitchell, B. J., 1990, A back-projection method for imaging large-scale lateral variations of Lg coda Q with applications to continental Africa: Geophys. J. Int., **100**, 161-181.

Yilmaz, O., 1987, Seismic data processing: Society of Exploration Geophysicists.

Zoeppritz, K., 1919, Uber reflexion und durchgang seismischer wellen durch unstetigkeitsflachen, Berlin, Uber Erdbebenewellen VII B, Nachrichten der Koniglichen Gesellschaft der Wissenschaften zu Gottingen, Math-Phys., K1, 57-84.

The vita has been removed from the digitized version of this document.

"This is nice, but just once I'd like to go out bowling!"